

Universal Coded Character Set
UCS

ISO/IEC JTC1/SC2/WG2 IRG N2743

Date: 2024-10-17

Source:	China
Author:	Tao Yang
Title:	Feedback on IRGN2730
Meeting:	IRG #63
Status:	Member Body Contribution
Actions required:	To be considered by IRG
Distribution:	IRG
Medium:	Electronic
Page:	2
Appendix:	None

Tianheng put forward a disunification suggestion about U+2335F (肩) and U+6B25 (吹).
天珩提出了关于 U+2335F 和 U+6B25 的拆分建议。

I agree in principle that Chinese characters composed of components with different meanings and glyph should be encoded separately.

Then I will elaborate on my views on these two characters.

我原则上同意由不同含义和字形的部件组成的汉字应该单独编码。

然后，我将详细阐述我对这两个字的看法。

1. Not necessary to overly rely on component unification rules

1.不必拘泥于部件认同别异规则

There is no absolute way to distinguish component confusion.

部件混同并无绝对的区分方式

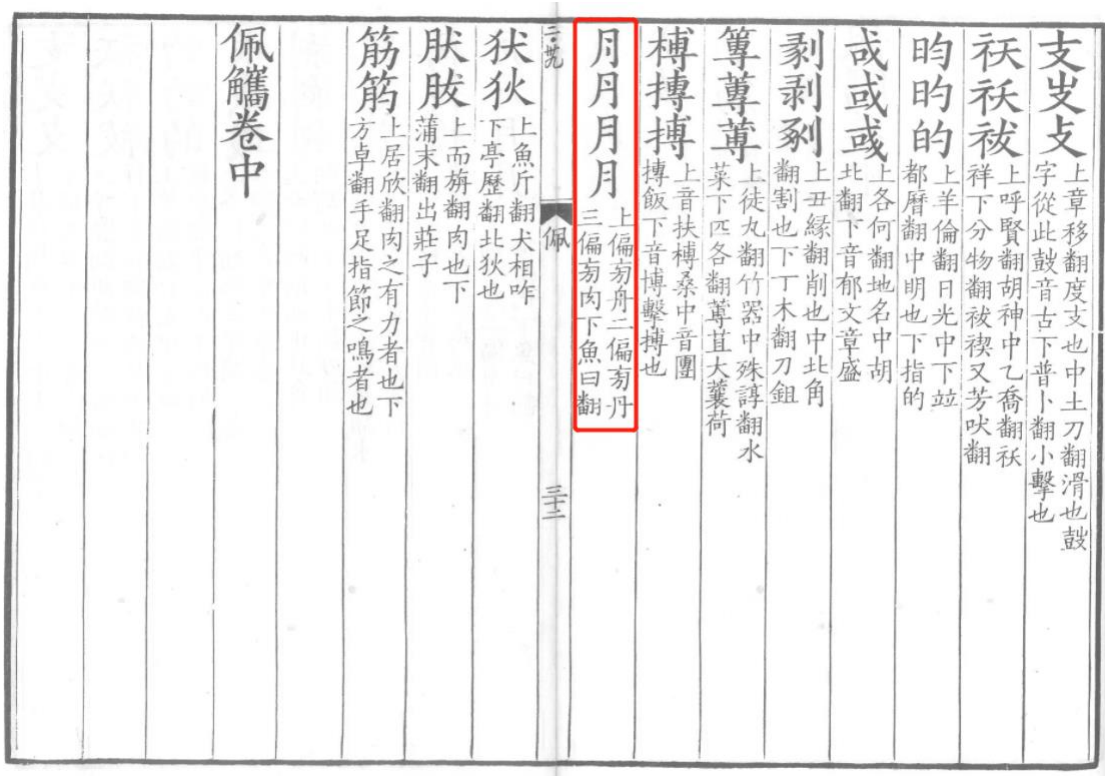
Components with specific pictographic and indicative meanings, such as "冎" used for "冒" and "平" used for "眾", retain their unique and rarely used original components finally due to their clear pictographic meanings, even if they are already been simplified and confused with other components. In the solid majority of cases, components naturally convergence and reduction during use, such as "土" and "士" used for "壘", "𠂇" and "丰" used for "契", and "人" and "入" used for "全", and so on.

具有特定的象形、指事含义的部件，如冎之于冒、平之于眾，即使出现了与其他部件的混同简化用法，但因其明确的象形意义而最终保留了特殊、罕用的原始构件。绝大多数情况下，部件都是在使用的过程中，自然形成了趋同和省减，如土士之于壘、丰之于契、人入之于全等等。

In later books on character analysis, there was a theory that two short horizontal strokes, with or without touching the right side stroke '冂', were used to distinguish between 'moon' and 'meat'.

However, this was limited to the rules of a certain type of character books and did not have analogies. When placed in a massive amount of literature, it would lose its foothold in the viewpoints, and the way of distinguishing components was not effective. Some rules were always stable, and some differences were extremely weak. Especially in the recent evolution of Chinese characters, '月' and '冂' were indistinguishable. Many differentiation rules gradually became invalid, making it difficult to distinguish problematic character shapes through general unification rules.

后来的字形辨析书籍中，有用两短横=接触不接触右边笔画冂来区分是月还是肉的说法，仅限于某一种字书的正形规则而不具备类推性，放在海量的文献中则失去了观点的立足点，区分部件的方式并不绝对，有些规则是一直稳定的，有些区别极为微弱，特别是在晚近的汉字演变中，月月无别，冂冂不分，许多区分规则逐渐失效，很难通过一般的认同别异观点去分辨疑难字字形。



2. U+2335F (冂)

This character originates from the fifth character **冂** in the "Moon" section of the Kangxi Dictionary, and the nineteenth character **冂** in the "Meat" section shares a similar shape. Hanyu Dazidian categorizes them both as homograph characters with two meanings, and adjusts the stroke shape in accordance with the mainland Chinese character form standard, resulting in the encoding issue.

这个字出自《康熙字典》月部第五字**冂**，而肉部第十九字有**冂**与之同形，《汉语大字典》

径将二者作为同形字收为一字两义，且按照大陆字形规范调整了笔形，编码问题遂由此而来。

𦘔 𦘕 說文·肉部
苦文

(一)kǎn

同“肯”。清高翔麟《說文字通·肉部》：“𦘔，古文作𦘔。今作肯。”清劉獻廷《廣陽雜記》卷四：“蓋其平日專用力于近侍之人，閣部九卿，全不留意，故滿漢大臣多少之，無人引見……而王輔臣又不肯受以金壁賄賂部臣，惟隨衆常朝而已。”

(二)wéi 《改併四聲篇海》引《搜真玉鏡》烏賄切。

吐。《改併四聲篇海·月部》引《搜真玉鏡》：“𦘔，吐也。”

Then let's analysis the record and usage of these two glyphs in the other dictionaries and literatures to determine whether is it proper to encode them separately.

那么我们来看看历史上的字书和文本对这两个康熙字头是怎么记录和应用的，适不适合将其中一个拆分出来独立编码：

2.1 Glyphs

2.1 字形

There are 4 glyphs in the actual usage of these 2 characters.

在实际应用中，这两个字头有如下 4 种字形。

𦘔 𦘕 𦘔 𦘕

2.2 Comparison of these 2 head characters in the same dictionaries

2.2 同一字书中存在两个字头的字形对照

Please see table 1, in the statistics of the character pronounces wei3 in the 月 section, "𦘔" appears 2 times, "𦘕" 7, "𦘔" 11, and "𦘕" 2.





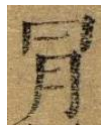
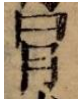



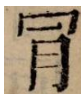

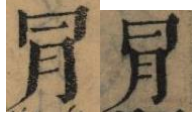
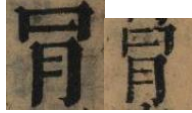







In the statistics of character pronounces Ken3 in the 肉 section, "𦘔" appears 11 times, "𦘕" 2, "𦘔" 5, and "𦘕" 18.






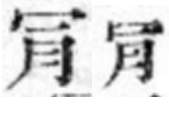

















请看表 1，月部 wei3 声字头的字形统计中，𦘔 出现 2 次，𦘕 出现 7 次，𦘔 出现 11 次，𦘕 出现 2 次，肉部 ken3 声字头的字形统计中，𦘔 出现 11 次，𦘕 出现 2 次，𦘔 出现 5 次，𦘕 出现 18 次。

According to the statistics, it is evident that there are 2 distinct head characters in 正字通/字彙/光霽字彙/鳳儀字彙/玉鼎字彙/增補文成字彙/元音統韻/康熙字典, which means they stably exist separately in the late antiquity dictionaries. Various dictionaries should have a thought about dividing them with obvious differences. Holistically speaking, all the distinction points are used in a chaotic manner, even in the same dictionary.

由以上统计可见，它们在《正字通》《字彙》《光霽字彙》《鳳儀字彙》《玉鼎字彙》《增補文成字彙》《元音統韻》《康熙字典》中分立两个字头，属于近古字书中稳定的字头字。各种字书确实有其将两个字头从字形上加以区分的目的。一种字书中，二者各有明确的区别点。而综合起来看，每个区别点实际上都处在混用状态，甚至同一种字书里都可能混淆两个字头的字形。

Table 1

wei3 (月部)	ken3 (肉部)
<p>正字通 1 () </p> <p>正字通 2 () </p>	<p>正字通 1 () </p> <p>正字通 3 () </p>
<p>光霽字彙 1 () </p>	<p>光霽字彙 1 () </p> <p>光霽字彙 2 () </p> <p>光霽字彙 3 () </p>
<p>鳳儀字彙 1 () </p>	<p>鳳儀字彙 1 () </p> <p>鳳儀字彙 2 () </p>
<p>玉鼎字彙 () () </p>	<p>玉鼎字彙 () </p>
<p>元音統韻 1 () </p>	<p>元音統韻 1 () </p> <p>元音統韻 2 () </p> <p>元音統韻 3 () </p>
<p>字彙 1 () () </p>	<p>字彙 1 () </p> <p>字彙 2 () </p>

<p>增補文成字彙 1 (J) </p> <p>增補文成字彙 2 (J) </p>	<p>增補文成字彙 1 (J) </p>
<p>康熙字典 1 (J) </p> <p>康熙字典 2 (J)  </p> <p>康熙字典 3 (J)  </p> <p>康熙字典 4 (J) </p>	<p>康熙字典 2 (J)  </p> <p>康熙字典 4 ()(J)   </p> <p>康熙字典 5 ()(J)()</p> <p>  </p> <p>康熙字典 6 ()()  </p> <p>康熙字典 7 () </p> <p>康熙字典 8 () </p> <p>康熙字典 9 () </p>
<p>集字繫聲 (J)(J)  </p>	<p>集字繫聲 (J)()  </p>
<p>五音集字 (J) </p>	<p>五音集字 () </p>
<p>辨字通考 1 () </p> <p>辨字通考 2 () </p>	<p>辨字通考 1 ()()  </p> <p>辨字通考 2 () </p>

2.3 The glyphs of these 2 characters appear in the dictionaries and literatures singly

2.3 两个字头在字书和文献中单见的情况

Please see the table 2, in the statistics of wei3 from 月 section, 冑 appears 0 times, 冑 0, 冑 4,

肩 10; in the statistics of ken3 from 肉 section, 肩 appears 0 times, 肩 0, 肩 7, 肩 28.
 请看表 2, 月部 wei3 声字头中, 肩出现 0 次, 肩出现 0 次, 肩出现 4 次, 肩出现 10 次, 肉部 ken3 声字头中, 肩出现 0 次, 肩出现 0 次, 肩出现 7 次, 肩出现 28 次。

Table 2

Wei3 (月部)	Ken3 (肉部)	
陈明卿太史考古詳訂遵韻海 篇朝宗 (J) 	四库全书考证 (J) 	藝文備覽 (I) (I) (I) 
重校經史海篇直音 (J) 	新體說文大字典 (I) 	所南文集 (J) 
重刊詳校篇海 (I)  重刊詳校篇海 (I) 	說文通訓定聲分部檢韻 (J) 	同音字辨 (J) (?) 
重刊訂正篇海 (I) 	說文解字 (I) 	古字彙編 (I) 
篇海類編 (I) 	說文長箋 (J) 	爾雅注疏校勘記 (I) 
精刻海若湯先生校訂音釋五 侯鯖字海 (I) 	說文通訓定聲 (I) 	經典釋文校勘記 (I) 
攷正玉堂字彙 (J) 	說文解字校錄 (I) 	經籍纂詁 (I) 
字貫 (I) 	說文正字 (I) 	元聲韻學大成 (I) 

合併字學篇韻便覽 (J)  合併字學篇韻便覽 (J) 	說文拈字 (I) 	韻話 (I) 
大明成化丁亥重刊改併五音類聚四聲篇 (I) 	說文重文 (I) 	八瓊室金石補正 (I) 
大明正德乙亥重刊改併五音類聚四聲篇海 (I) 	說文通檢 (I) 	五山耆舊今集初刊 (I) *19 
新刻辨疑正韻同文玉海 (I) 	說文字通 (I) 	寧都三魏全集 (I) (I) 
	識字略 (I) 	惜陰軒叢書·兩山墨談 (I) *3 
		二十二子·呂氏春秋 

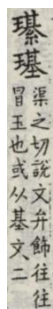
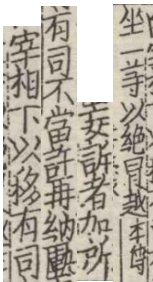
2.4 冒 is also usually used as the variant of 冒, slightly different from 冒


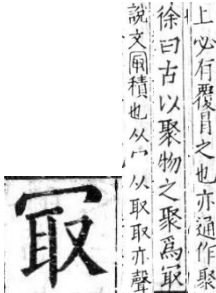
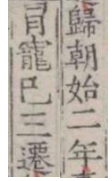
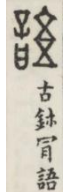
2.4 冒常常作为冒的异体字冒使用

Please see table 3.

请看表 3。

Table 3

冒		
类篇 	太學新增合璧聯球聲律萬卷菁華 	北堂書鈔

		
<p>古今韻會舉要小補</p> 	<p>白氏文集</p> 	<p>說文古籀補 ()</p> 

2.5 Summary

2.5 小结

We can be surprised to see that, 冒冒, which were distinguished in the earlier dictionaries, were disappeared in the actual usage, and the actual use of 冒冒 has not produced an equal proportion of glyph differences according to different meanings.

我们惊奇的发现，较早的字书中用来作为字头的冒冒在实际使用中完全消失了，而在用的冒冒也并未按照意义产生比例区别。

That means, the chaotic usage of them have been a common phenomenon for a long time.

这也就是说，实际使用中二者的混用由来已久，且成为了普遍现象。

The separation of 2 head characters in one dictionary does have considerations for distinguish them, but really difficult for the common people as rarely-used characters.

作为字书而言，分立两个字头，确有分立两个字形的考量，但作为一般用字而言，这两个字本就生僻，区分二者颇有难度，实际书写中，笔画从便从简，字形逐渐向冒冒聚合。

3 U+6B25 (吹)

3.1 Glyphs

3.1 字形

There are 7 different glyphs in the practical use, of which the first 3 are 吹 actually, the 4th is the one that Tianheng considers to be distinguished, the last 3 are not within the scope of discrimination.

这组字符实际使用中有 7 种字形，其中前 3 种实际都是吹，第 4 种则是天珩考虑需要辨析的吹[𠂔曰欠]字形，后 3 种不在辨析范围内。

吹 吹 吹 吹 𠂔 𠂔 𠂔

3.2 The situation where these 2 characters appear in dictionaries without separated head characters

3.2 未分立字头的字书中出现这两个字形的情况

Please see table 4, in the series of Shuowen dictionaries, a lot of the glyphs were written as 吹 with quote 字从曰 only, while the others written as 吹[𠂔曰欠].

请看表 4，说文系统的字书中，不少明确将字形写成吹[𠂔曰欠]，而多数则只在字条中表述，该字从曰，但字形仍为吹。

Thus, this kind of dictionaries don't view them as different characters, both aim to the same head character 𠂔. The unique difference is some dictionaries design the glyph as 吹[𠂔曰欠] according to the meaning strictly, and the others didn't aware of the point.

因此未分立字头的字书，吹吹都指向同一个字头 𠂔，未将这两个字形视为不同的字，只是有的严格按照释义书写成吹[𠂔曰欠]，有的没有意识到部件差异，减省书写为吹。

Table 4

𠂔	
𠂔曰欠	𠂔曰欠
說文解字 1 (曰) 吹	說文解字 6 (曰) 吹
說文解字 2 (曰) 吹	
說文解字 3 (曰) 吹	
說文解字 4 (曰) *2 吹吹	
說文解字 5 (曰) 吹	
說文解字繫傳 1 (曰) 吹	說文解字注 1 (曰) *18 吹
說文解字繫傳 2 (曰) 吹	說文解字注 2 (曰) *8 吹

說文解字繫傳 3 (日) 吹 說文解字繫傳 4 (日)	
說文二徐異訓辨 (日) 吹	說文句讀 1 (日) *4 吹 說文句讀 2 (日) *4 吹
說文字原集注 (日) 吹	說文解字篆韻譜 (日) 吹
說文字原考略 (日) 吹	重刊許氏說文解字五音韻譜 (日) 吹
說文理董後編 (日) 吹	說文解字補義 (日) 吹
集韻 (日) *4 吹 吹	說文字母集解 (日) *55 吹 吹
濂亭文集 (日) 吹	十三經音略 (日) *2 吹 吹 (日) *2 吹 吹
文字蒙求廣義 (日) *2 吹	汲古閣說文訂 (日) *14 吹
說文解字義證 (日) *14 吹	說文解字注箋 (日) *7 吹 (日) *5 吹
說文通訓定聲 (日) *6 吹 吹 (日) 日吹 吹	文字蒙求 1 (日) 吹 吹 文字蒙求 2 (日) 吹 吹
段注說文解字斟誤 (日) 吹	說文解字校錄 (日) 吹
惠氏讀說文記 (日) *3 吹	說文校議議 1 (日) *2 吹 說文校議議 2 (日) *2 吹
頤志齋叢書 (日) *3 吹	說文解字羣經正字

	(日)*7 吹吹
說文解字注匡謬 1 (日)*2 吹	說文引經異字 (日)*6 吹
說文解字注匡謬 2 (日)*3 吹	
說文解字斟詮 (日)*3 吹	說文正字 (日) 吹
說文辨字正俗 (日) 吹	說文蠡箋 (日)*7 吹
說文拈字 (日)*7 吹	說文類解 (日)*3 吹
說文經典異字釋 1 (日)*4 吹	
說文經典異字釋 2 (日)*4 吹	
說文字通 (日)*11 吹	
說文測議 (日) 吹	
說文長箋 (日)*16 吹吹	
龍龕手鏡 (日)*2 吹	
班馬字類 (日)*4 吹吹	
班馬字類 (日)*4 吹吹	

3.3 The situation where these 2 glyphs separated to 2 head characters in the dictionaries

分立字头的字书中两种字形的存在情况

Please see table 5, 3 new pronunciations and meanings appear in the dictionaries which separated 2 head characters.

请看表 5，分立字头的字书中，出现了三种新的音义：

The 1st one, huan1 歡, this is the origin of the pronunciation 音歡 in 大漢和辭典 quoted by IRGN2730.

第 1 种，huan1 歡，这是 IRGN2730 引用的《大漢和辭典》中读音“音歡”的来历。



The 2nd one, yi4/yu4 詞/辭, a mistake copy from the original text of 說文解字.

第 2 种，yi4/yu4 詞/辭，《说文解字》中的释义是“詮詞也。从欠从日，日亦聲。”几种字

书释义时省作“词也”“辞也”，发生了意义偏转。

The 3rd one, xin1 昕, which means dawn. This is the origin of the meaning sóm(清晨) in 字典字喃引解 quoted by IRGN2730.

第3种, xin1 昕, 意为天亮。这是 IRGN2730 引用“字典字喃引解”中“só m(清晨)”的来历。

There is glyph distinction between the separated head characters. The one pronounces yu4 was written as  with the last stroke changed to 丿, while the new pronounce huan1 歡 was written as 日 in 新校經史海篇直音. 重刊詳校篇海 distinguish them by component 日 and 日. And both of these 2 characters were written as  in 精刻海若湯先生校訂音釋五侯鯖字海. So this kind of distinction is chaotic and lacks practical significance.



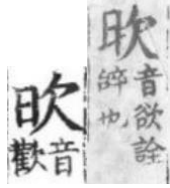
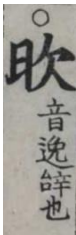

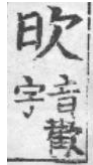
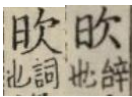


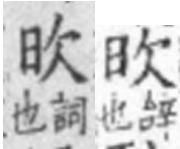
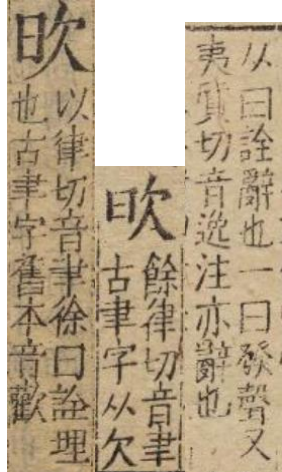

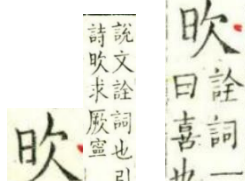
分立字头也有一定的字形区别, 《新校經史海篇直音》表示原义的 yu4 写作提笔 , 表示新增义 huan1 写作日, 《重刊詳校篇海》是用日 and 日来做区分, 《精刻海若湯先生校訂音釋五侯鯖字海》则两个都作 , 所以这种字头区别混乱无章, 没有实际意义。

Table 5

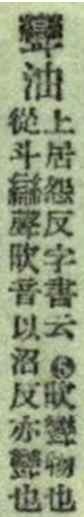
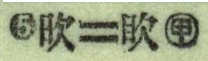
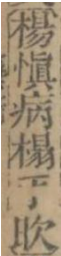


新音义		
huan1 歡	yi4/yu4 詞/辭	xin1 昕
新校經史海篇直音 	龍龕手鏡 	陈明卿太史考古詳訂遵韻海篇朝宗 
大明正德乙亥重刊改併五音類聚四聲篇海 	鉅宋廣韻 	新刻辨疑正韻同文玉海 
重校經史海篇直音 	廣韻 	
重刊詳校篇海	重修廣韻	

 <p>吹 也古聿字舊本音歡</p> <p>吹 餘律切音聿古聿字从欠</p> <p>吹 從口詮辭也一曰發聲又夷質切音逸注亦辭也</p>	 <p>吹 也詞</p> <p>吹 也辭</p>	
<p>精刻海若湯先生校訂音釋五侯鯖字海</p>  <p>吹 音欲詮辭也</p>		
<p>字彙補</p>  <p>吹 許官切音歡見篇韻與吹字同吹从日音聿者不</p>		
<p>大明成化丁亥重刊改併五音類聚四聲篇海</p>  <p>吹 字音歡</p>		
<p>集韻</p>  <p>吹 說文詮詞也一曰喜也詩吹求厥室引</p>		

3.4 The other usage in the literatures

3.4 在文献中出现的其他用法

All of them are wrong hand-writing or mistakenly used.
 均为误写误用。

chui1 吹	yao3 汲水	yi4/yu4 詞/辭
少年自然科学丛书:第九编 我們地球如果遇着這風, 怕便會吹成微塵。 那要一瞬間。 吹	續一切經音義   吹	通俗編  病榻手吹
戏学汇考  吹		
歷代小史  吹		

3.5 Summary

3.5 小结

In the series of Shuowen dictionaries, distinguish or not between 日 and 日 coexist. [日欠] is based on the original meaning, while [日欠] for convenience.

在说文系字书中，区分和不区分日欠和日的情况同时存在，写作“日欠”是基于字源字理，写作“日欠”则是便宜行事。

In the other kinds of dictionaries, the original meaning doesn't distinguish 日 and 日. When a new semantic term appears, there are cases where the head characters are separated, but few of the glyphs are distinguished.

而其他字书中，作为本义并不为区分日欠和日；出现新义项时，存在分立字头的情况，然而区

别字形的情况很少。

There is a accurate discourse on 吹 in 說文廣義 for reference:

《說文廣義》对“吹”有一段非常准确的论述，供大家参考：

吹/聿/遁 share the same pronunciation 余律切, but distinguished by meanings……Those who continue to study 詩經 and 尚書 generally use 聿 and 遁 instead of 吹, rendering the character '吹' obsolete……

吹、聿、遁，三字同音余律切，而義別。聿，筆也，楚謂之聿，秦謂之筆。遁，回避也，姦不任事，謂之回遁。唯吹从日从欠；日，言也，欠，亦言也，故爲發語詞，與日、爰、于通用。傳寫詩、書者概以聿、遁代吹，吹字遂廢。凡此類，皆因經師注音於本字之下，習者不審，以注音爲本字，口授差舛，無有知其別者，乃至讀聿、遁爲律音，莫有改正之者。

4 Consensus

4.总结

For the effect of splitting the characters into different codepoints, especially the URO characters, is the most important issue for consideration. If it does, the splitting should not be undertaken rashly.

No matter of which meaning the head character is, the accurate glyphs of U+6B25 (吹) has already undergone a trend towards simplification and unification in its form, and a large number of digital achievements have already been made with it, and cannot be ruled out that it has already been used in personal names.

Therefore, U+6B25 should retain the 吹 [𠄎日欠] form. Whether to split [日欠] into a new code point requires the researchers and publishers of 說文解字 to propose the need. Currently, the significance of distinguishing it is not clear.

作为 URO 的字符，是否将不同源的字形拆分到不同码位，首先应当考虑拆分后是否影响各地区间的数据交换，如果影响，则不应贸然拆分。U+6B25 (吹)不论作为哪个意义的字头，字形已经趋于简化统一，且已经形成了大量数字化成果，且不排除已经用于人名，因此 U+6B25 应保留吹[𠄎日欠]字形，是否拆分[日欠]至一个新的码位，则需要《说文解字》的研究者和出版者提出需求。目前来看，对其区分的意义并不明显。

U+2335F (𠄎) with a small seal script origin, is already quite commonly used as the variant of 肯. However, as an onomatopoeic character for spoken "吐", it is not familiar to people, the theoretical basis for its glyph is not clear enough to determine the differences between 𠄎 𠄎 𠄎 𠄎.

There are indeed some rules for the perspective of glyph statistics. According to the comparison of the two head characters in the same word book, 𠄎 𠄎 appear 18 times in total, 𠄎 𠄎 appear 4 for using as wei3 with radical 月, while 𠄎 𠄎 appear 29 times in total, 𠄎 𠄎 appear 7 for using as ken3 with radical 肉. Wei3 tends to the glyph with component 月, and ken3 prefer 𠄎.

However, the PRC convention doesn't distinguish 月 and 𠄎, and all the lower part components are all shaped as 𠄎.

Thus if it's really necessary to split them, 𠄎 should be retained as the glyph of ken3 on the

code point U+2335F, and separate 𠂔 to a new code point as the glyph of wei3.

U+2335F (𠂔)作为肯的异体字已经较为常见，且有小篆字形源头，而作为口语“吐”的拟声词则不为人熟悉，字形理据也不够明确，并不能从理据角度判定字形差异点；从字形统计角度确实有一些规律，根据分列两个字头的字书对比情况来看，月部 wei3 声字头，𠂔 𠂔合计出现 18 次，𠂔 𠂔合计出现 4 次，肉部 ken3 声字头中，𠂔 𠂔合计出现 29 次，𠂔 𠂔合计出现 7 次，wei3 更倾向于从月的字形，ken3 更倾向于从𠂔，但中国大陆字形规范已经不区别月𠂔，且在下的部件一律定形为𠂔，因此如果确实需要拆分，应在 U+2335F 保留𠂔形作为“肯”的异体，而分出𠂔作为 wei3 的字形。

Strictly speaking, both sets of characters have the possibility of distinguishing the codepoints. Considering the risk, U+6B25 (𠂔) should not be split rashly. If necessary for textual research and dictionary publishing, the 𠂔日欠 can be split into other code point, while the original code position retains the 𠂔日欠; U+2335F (𠂔) can be considered for splitting, with the original code position retaining the character 𠂔(ken3) and a new code position serving as 𠂔 (wei3).

两组字头严格来说都有区分字形的可能性，考虑风险问题，U+6B25 (𠂔)不应贸然拆分，文字学研究和字书出版如有需要，可将𠂔日欠拆分到其他码位，原码位保留𠂔日欠；U+2335F (𠂔)可以考虑拆分，原码位保留𠂔形（ken3），新码位作为𠂔 wei3。

End of the Document.