

IRG N2174 Part 2 A

Title: Code-point for Full-width Separator Dot (音界號 / 間隔號)

Source: Henry Chan

Type: Individual Contribution to IRG #47

Date: 2016/10/12

Introduction

There are currently two code-points used for the separator dot in Chinese punctuation: U+00B7 Middle Dot and U+2027 Hyphenation Point. However, the fullwidth/halfwidth property of these two codepoints are not defined and implemented consistently. These two codepoints are also used in other Latin-based scripts, which is expected to be proportional in those contexts.

Description

For PRC, GB2312 maps the separator dot to U+00B7. For SimSun and Kaiti, U+00B7 is rendered as full-width.

For Taiwan, default operating system fonts do not handle this consistently. U+2027 should be used according to CNS11643, and should be full-width, but 新細明體 displays as proportional, while 標楷體 displays as full-width.

For HK, the big-5 mapping should be to U+00B7, however the neither MSungHK nor LiSung render the code-point as full-width.

In these case of mixed script use (such as Han/Eng), applications must implement complex language detection combined with specific fonts to show the full-width or half-width version in the correct context.

Action Item

In order to ensure that full-width and half-width separator dots are displayed correctly, a different codepoint should be used for the full-width version, as with other CJK Punctuation. It is suggested that the existing "U+30FB Katakana Middle Dot" be unilaterally used for this purpose.

This concerns the mapping of GB2312 0xA1A4, Big5 0xA145, and CNS 1-2126.

Supplementary Information

The request to use U+30FB was also raised at Unicode PRI 316 (<http://www.unicode.org/review/pri316/>) by Bobby Tung from Taiwan.

(Sometimes the U+FF0E full-width full stop is also used for the separator dot, however, it should be considered as misuse due to its name.)

The right hand side for 秋 (穉, 穉, 穉, 𪛗, 穉, 穉, 𪛗, 𪛗, 𪛗) and the like have always been thought to be 龜. However, recent oracle bone script studies indicate that the right hand side is not a turtle but likely locusts instead.

The confusion has been evidenced inside shuowen where the ancient form was erroneously drawn with a turtle head instead of a locust head.



Recently some scholars have “modernized” the oracle bone form into :

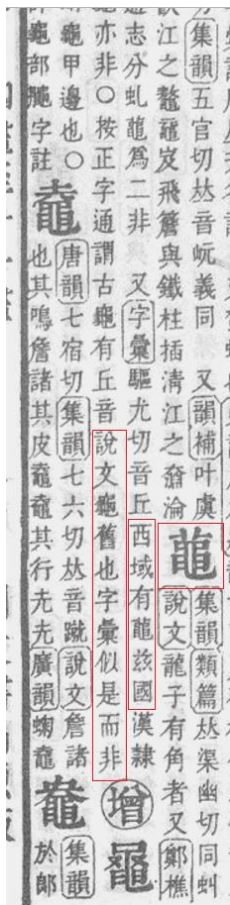
《甲骨文字典》頁783：徐中舒主編，四川辭書出版社，1989

龜

<p>甲 骨 文 字 典 卷 七</p> <p>七 八 三</p>	<p>一、紀時名詞，春秋之秋。</p> <p>二、帝秋、告秋、秬秋，皆有關秋時之祭祀，或謂為有關蝗神之祀。</p>	<p>〔釋義〕</p> <p>一、紀時名詞，春秋之秋。</p> <p>二、帝秋、告秋、秬秋，皆有關秋時之祭祀，或謂為有關蝗神之祀。</p>	<p>〔解字〕唐蘭謂象龜屬之動物，即萬象名義之龜字，又疑為說文之龜字，為水蟲之一種，借為春秋之秋。殷虛文字記按當以釋龜為是。後世謠為龜，更增禾旁為穉，為秋之異體。見於《集韻》。又，卜辭亦借穉為龜，見卷十火部燧字說解。此即說文之秋字籀文所本。或謂穉等形象蝗形，為蝗之初文，於卜辭辭例亦可順釋。按其說可參。《說文》：「秋，禾穀熟也。从禾，夔省聲。」</p>	<p>龜 一期甲二三五三 龜 一期前五二五二 龜 一期後下二四 龜 一期合二八 龜 一期存二九二</p> <p>龜 一期林二八二 龜 一期林二八四</p> <p>龜 一期林二八二 龜 一期林二八四</p> <p>龜 一期林二八二 龜 一期林二八四</p> <p>龜 一期林二八二 龜 一期林二八四</p>
---	---	---	--	--


(remarks in blue)

At the same time, 龍 is also encoded at U+2A4D4.





Which also gives a pronunciation similar to 秋 and also quotes <<字彙>> to say that it is the correct form of the 龜 in “龜茲”. (though author of Kangxi Dictionary disputes this by referencing shuowen, which has been proved wrong.)

Possible Action items:

A1) Encode  as a new character; or

A2) Unify it with  (U+2A4D4); or

A3) Do not encode

B1) Allow changing of all characters with shape  but with etymology of 秋 to use  instead (i.e. addition to UCV)

B2) Do not add to UCV and encode the new variants with  only when they appear in other (reputable) text.

-- End of Document