

Using ZWJ in the encoded representation of Brahmi numbers

Shriramana Sharma, jamadagni-at-gmail-dot-com, India

2010-Sep-14

The Brahmi proposal L2/07-342 in sec 3.5 (pp 6-7) indicates that in Brahmi:

"Multiples of 100 and of 1000 are expressed multiplicatively, with the multiplier following and forming a ligature with 100 or 1000"

and goes on:

"we suggest that these ligatures be encoded with ZERO WIDTH JOINER (200D)."

later providing a sample:

| | |
|--------------|-----------------------|
| 𑀓𑀕 | 𑀓𑀕𑀕 |
| 100 4 | 100-4 |
| (= 104) | (= 400) |
| 11063, 11054 | 11063, 200D, 11054 |

in which the *unligated* form of the symbol of number 100 followed by that of number 4 meaning $100 + 4 = 104$ is represented by the sequence BRAHMI NUMBER HUNDRED + BRAHMI NUMBER FOUR whereas the *ligated* form of the same symbol of number 100 followed by that of number 4 meaning $100 \times 4 = 400$ is represented by the sequence BRAHMI NUMBER HUNDRED + ZWJ + BRAHMI NUMBER FOUR.

However this would be against Unicode principles because it makes the joiner semantically significant. Since the joiners are ignored in text search and collation, searches for 104 would turn up 400 and so on which is obviously not desired.

As a better way to make the semantic distinction between the ligated and unligated forms, I suggest using 2062 INVISIBLE TIMES. While BRAHMI NUMBER HUNDRED + BRAHMI NUMBER FOUR would naturally be presented as the unligated sequence of two symbols and clearly mean at both the visual and encoding levels $100 + 4 = 104$, we can prescribe BRAHMI NUMBER HUNDRED + INVISIBLE TIMES + BRAHMI NUMBER FOUR to be represented as the ligature and hence clearly mean at both the visual and encoding levels $100 \times 4 = 400$. Thus:

BRAHMI NUMBER HUNDRED + BRAHMI NUMBER FOUR \rightarrow No ligature = 104

BRAHMI NUMBER HUNDRED + INVISIBLE TIMES + BRAHMI NUMBER FOUR \rightarrow Ligature = 400

-O-O-O-