L2/26-035

On Unicode Character Database Update of U+1BBA

Febri Muhammad Nasrullah < niomi13@pm.me >
October 15th, 2025

Introduction

This proposal requests an update in the Unicode Character Database of the Unicode character U+1BBA SUNDANESE AVAGRAHA.

Although Unicode character properties are generally stable, the Character Encoding Stability Policy allows corrections where evidence shows that a property was assigned based on incomplete understanding. Such is the case with U+1BBA, which was encoded in 2009 (Everson, L2/09-372) based on limited data about Sundanese orthography. As a result, it was categorized under *Lo* (*Other Letter*), based on the assumption that it functioned similarly to the *avagraha* in other Indic scripts.

Subsequent research—particularly Nurwansah (2021, pp. 4-7) —clarifies that U+1BBA functions not as an *avagraha* but as a gemination marker attached to a preceding consonant. It cannot appear word-initially, can be followed by vowel signs, and marks consonant doubling rather than elision. Therefore, its current classification as *Lo* doesn't reflect its actual orthographic behavior in historical Sundanese manuscripts.

Reclassifying U+1BBA as *Mc* will improve rendering consistency across shaping engines (e.g., HarfBuzz, Uniscribe), enable accurate syllable segmentation in text processing, and bring Sundanese script behavior in line with comparable Indic gemination marks. This correction will not destabilize existing character identities or mappings, as it preserves the code point and name, modifying only property semantics.

Requested Changes

1. UnicodeData.txt
   `General_Category: Lo → Mc`
2. IndicPositionalCategory.txt
   Add: `Indic_Positional_Category: Right`
3. IndicSyllabicCategory.txt
   `Indic_Syllabic_Category: Avagraha → Gemination_Mark`
4. LineBreak.txt
   `Line_Break: AL → CM`
5. NamesList.txt
   Add comment: `% SUNDANESE GEMINATION MARK`
6. NameAliases.txt
   Add: `correction`
7. PropList.txt
   `Other_Alphabetic → Diacritic`

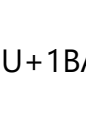
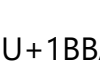
Rendering Behavior and Expected Changes

Tests were conducted across shaping engines and operating systems using system-embedded Noto Sans Sundanese/Sans Serif Collection and RL Hanjuang (custom OT font). The sequence <U+1B8A, U+1BBA, U+1BA6> was used to observe contextual placement of U+1BBA relative to base and vowel signs.

Shaping Engine / Platform	*Current*		*Expected*
	System Fonts	*RL Hanjuang*	
HarfBuzz (Mozilla Firefox)			
HarfBuzz (LibreOffice Writer)			
HarfBuzz (Android 15)			
USE (Office 365 Word)			
USE (Notepad)			
CoreText (macOS)			

RL Hanjuang demonstrates consistent rendering behavior across all tested platforms and shaping engines, in contrast to system-embedded fonts, which exhibit inconsistencies. Notably, the system font used in Notepad represents the latest version, yet its rendering output aligns closely with Android 15, indicating that the underlying shaping logic remains unchanged. Since U+1BBA is currently classified as a *Letter, Other (Lo)* rather than a *Mark, Spacing Combining (Mc)*, the rendered output deviates from the expected contextual arrangement.

The table below illustrates a hypothetical scenario in which U+1BBA is reclassified as *Mc*. The first row provides a comparison to U+1BA7, which is already categorized as *Mc*.

Sequence	System Fonts	*RL Hanjuang*
<U+1B8A, U+1BA7, U+1B80>		
<U+1B8A, U+1BBA, U+1BA4>		

<U+1B8A, U+1BBA, U+1BA5>		
<U+1B8A, U+1BBA, U+1BA6>	 	
<U+1B8A, U+1BBA, U+1BA7>		
<U+1B8A, U+1BBA, U+1BA8>		
<U+1B8A, U+1BBA, U+1BA9>		

In the first row, for instance, system fonts (except on Notepad and Android 15) correctly position U+1B80 above the base letter (*Lo*), consistent with historical manuscript evidence (Figure 1). This is because, similar to the Javanese script, *Mn* characters in Sundanese are expected to attach above or below the base consonant, not above *Mc* characters. RL Hanjuang follows this principle as well.

However, unlike Javanese, the Sundanese script lacks a dedicated shaping engine. Consequently, without OpenType-level intervention, *Mn* marks are erroneously positioned above *Mc* characters—as seen in Notepad and Android 15. The required intervention involves OpenType reordering rules that compel the sequence <U+1B8A, U+1BA7, U+1B80> to be rendered as <U+1B8A, U+1B80, U+1BA7>. RL Hanjuang applies this reordering mechanism and extends it to encompass all possible combinations of U+1BBA + vowel signs.

System fonts, however, appear not to implement this rule extension and continue to treat U+1BBA as *Lo*. The issue becomes particularly evident in the sequence <U+1B8A, U+1BBA, U+1BA6>, where U+1BA6 — which should appear before the base consonant (U+1B8A) — is instead rendered before U+1BBA, as shown in the table. This misplacement occurs because both U+1B8A and U+1BBA currently share the *Lo* property.

Therefore, U+1BBA should be reclassified as *Mc* to align with both expected OpenType behaviour and historical orthographic evidence (see Figure 2). Furthermore, the author requests that the Sundanese shaping engine adopt the same Indic reordering logic as used for the Javanese script, since Old Sundanese follows the same structural and phonographic principles. In particular, *Mn* signs should attach above or below the base consonant, not after *Mc* marks—reflecting the established Indic rendering model implemented in Javanese.

Figures

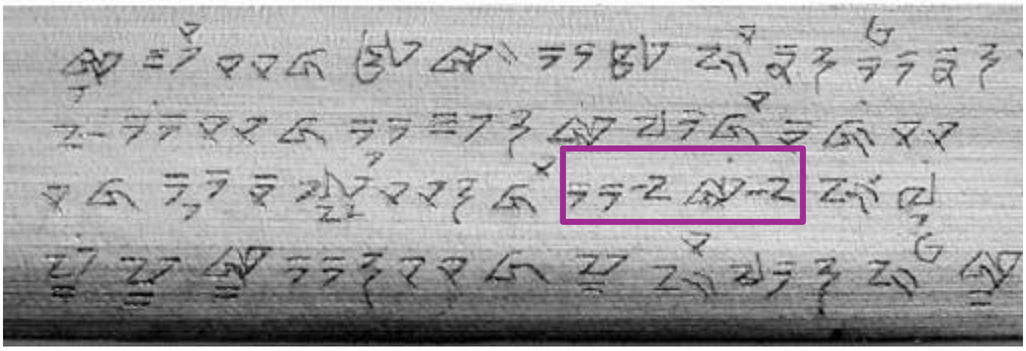


Figure 1. L 407 manuscript – example of sequence <U+1B98, U+1BA7, U+1B80> showing expected Mn *placement. Purple box:* ᮒᮒᮜᮘᮧᮀᮜ.

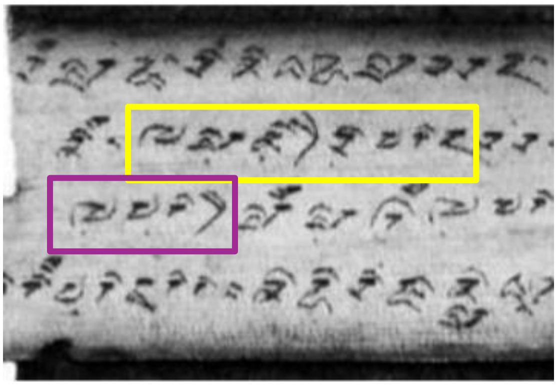


Figure 2. [L 624](#) 7r. – examples of <base letter, U+1BBA, U+1BA5> confirming direct vowel attachment to the base letter. Yellow box: ᮕᮓᮺᮥᮔᮺᮥᮔᮺᮥᮕᮺᮥᮒᮜ. *Purple box:* ᮕᮺᮥᮕᮺᮥᮒ.

References

Nurwansah, I. (2021, September 28). *Wrong Identities of Three Historical Sundanese Character*. Retrieved from Unicode: https://www.unicode.org/L2/L2021/21221-three-sundanese-chars.pdf

Everson, M. (2009, September 05). *Proposal for encoding additional Sundanese characters for Old Sundanese in the UCS.* Retrieved from Unicode: https://www.unicode.org/L2/L2009/09251r-n3666r-sundanese.pdf