Contributors:

Dr. Balaram Prasain

Dr. Laxmi Prasad Khatiwada

Mr. Hempal Shrestha

Mr. Pawan Chitrakar

Mr. Bhim Dhwoj Shrestha

Mr. Dev Das Manandhar

Dr. Bal Krishna Bal

Inputs from the Devanagari Nepali and newar Language Community for the nBGP Report

# Code point repertoire for the Devanagari Nepali and Newar languages

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Sr. No.** | **Unicode Code Point** | **Character** | **Character Name** | **Unicode General Category (gc)** | **Indic Syllabic Category** | **Reference** |
| 1. | 0901 | ँ | DEVANAGARI SIGN CANDRABINDU | Mn | Chandrabindu / Anunasika |  |
| 2. | 0902 | ं | DEVANAGARI SIGN ANUSVARA | Mn | Anusvara / Bindu |  |
| 3. | 0903 | ः | DEVANAGARI SIGN VISARGA | Mc | Visarga |  |
| 4. | 0905 | अ | DEVANAGARI LETTER A | Lo | Vowel |  |
| 5. | 0906 | आ | DEVANAGARI LETTER AA | Lo | Vowel |  |
| 6. | 0907 | इ | DEVANAGARI LETTER I | Lo | Vowel |  |
| 7. | 0908 | ई | DEVANAGARI LETTER II | Lo | Vowel |  |
| 8. | 0909 | उ | DEVANAGARI LETTER U | Lo | Vowel |  |
| 9. | 090A | ऊ | DEVANAGARI LETTER UU | Lo | Vowel |  |
| 10. | 090B | ऋ | DEVANAGARI LETTER VOCALIC R | Lo | Vowel |  |
| 11. | 090F | ए | DEVANAGARI LETTER E | Lo | Vowel |  |
| 12. | 0910 | ऐ | DEVANAGARI LETTER AI | Lo | Vowel |  |
| 13. | 0913 | ओ | DEVANAGARI LETTER O | Lo | Vowel |  |
| 14. | 0914 | औ | DEVANAGARI LETTER AU | Lo | Vowel |  |
| 15. | 0915 | क | DEVANAGARI LETTER KA | Lo | Consonant |  |
| 16. | 0916 | ख | DEVANAGARI LETTER KHA | Lo | Consonant |  |
| 17. | 0917 | ग | DEVANAGARI LETTER GA | Lo | Consonant |  |
| 18. | 0918 | घ | DEVANAGARI LETTER GHA | Lo | Consonant |  |
| 19. | 0919 | ङ | DEVANAGARI LETTER NGA | Lo | Consonant |  |
| 20. | 091A | च | DEVANAGARI LETTER CA | Lo | Consonant |  |
| 21. | 091B | छ | DEVANAGARI LETTER CHA | Lo | Consonant |  |
| 22. | 091C | ज | DEVANAGARI LETTER JA | Lo | Consonant |  |
| 23. | 091D | झ | DEVANAGARI LETTER JHA | Lo | Consonant |  |
| 24. | 091E | ञ | DEVANAGARI LETTER NYA | Lo | Consonant |  |
| 25. | 091F | ट | DEVANAGARI LETTER TTA | Lo | Consonant |  |
| 26. | 0920 | ठ | DEVANAGARI LETTER TTHA | Lo | Consonant |  |
| 27. | 0921 | ड | DEVANAGARI LETTER DDA | Lo | Consonant |  |
| 28. | 0922 | ढ | DEVANAGARI LETTER DDHA | Lo | Consonant |  |
| 29. | 0923 | ण | DEVANAGARI LETTER NNA | Lo | Consonant |  |
| 30. | 0924 | त | DEVANAGARI LETTER TA | Lo | Consonant |  |
| 31. | 0925 | थ | DEVANAGARI LETTER THA | Lo | Consonant |  |
| 32. | 0926 | द | DEVANAGARI LETTER DA | Lo | Consonant |  |
| 33. | 0927 | ध | DEVANAGARI LETTER DHA | Lo | Consonant |  |
| 34. | 0928 | न | DEVANAGARI LETTER NA | Lo | Consonant |  |
| 35. | 092A | प | DEVANAGARI LETTER PA | Lo | Consonant |  |
| 36. | 092B | फ | DEVANAGARI LETTER PHA | Lo | Consonant |  |
| 37. | 092C | ब | DEVANAGARI LETTER BA | Lo | Consonant |  |
| 38. | 092D | भ | DEVANAGARI LETTER BHA | Lo | Consonant |  |
| 39. | 092E | म | DEVANAGARI LETTER MA | Lo | Consonant |  |
| 40. | 092F | य | DEVANAGARI LETTER YA | Lo | Consonant |  |
| 41. | 0930 | र | DEVANAGARI LETTER RA | Lo | Consonant |  |
| 42. | 0931 | ऱ | DEVANAGARI LETTER RRA | Lo | Consonant |  |
| 43. | 0932 | ल | DEVANAGARI LETTER LA | Lo | Consonant |  |
| 44. | 0935 | व | DEVANAGARI LETTER VA | Lo | Consonant |  |
| 45. | 0936 | श | DEVANAGARI LETTER SHA | Lo | Consonant |  |
| 46. | 0937 | ष | DEVANAGARI LETTER SSA | Lo | Consonant |  |
| 47. | 0938 | स | DEVANAGARI LETTER SA | Lo | Consonant |  |
| 48. | 0939 | ह | DEVANAGARI LETTER HA | Lo | Consonant |  |
| 49. | 093E | ा | DEVANAGARI VOWEL SIGN AA | Mc | Matra |  |
| 50. | 093F | ि | DEVANAGARI VOWEL SIGN I | Mc | Matra |  |
| 51. | 0940 | ी | DEVANAGARI VOWEL SIGN II | Mc | Matra |  |
| 52. | 0941 | ु | DEVANAGARI VOWEL SIGN U | Mn | Matra |  |
| 53. | 0942 | ू | DEVANAGARI VOWEL SIGN UU | Mn | Matra |  |
| 54. | 0943 | ृ | DEVANAGARI VOWEL SIGN VOCALIC R | Mn | Matra |  |
| 55. | 0947 | े | DEVANAGARI VOWEL SIGN E | Mn | Matra |  |
| 56. | 0948 | ै | DEVANAGARI VOWEL SIGN AI | Mn | Matra |  |
| 57. | 094B | ो | DEVANAGARI VOWEL SIGN O | Mc | Matra |  |
| 58. | 094C | ौ | DEVANAGARI VOWEL SIGN AU | Mc | Matra |  |
| 59. | 094D | ् | DEVANAGARI SIGN VIRAMA | Mn | Halant / Virama |  |

Note: The character ऱ should be included in the MSR for Devanagari Nepali/Newar if it is not used as an isolated character but just for the purpose of producing eyelash ‘ra’.

This way, we have settled with 59 codepoints as the MSR for Devanagari Nepali and Newar.

# Composite characters – Confusingly similar shapes

It was noted that we don’t have any issues with the single characters like घ, ध, म, भ. All of these characters appear to be distinct in shapes. But there are issues with composite characters which are listed below:

|  |  |  |
| --- | --- | --- |
| द्ग  U+0926 U+094D  U+O917 | द्र  U+0926 U+094D  U+0930 | द्न  U+0926 U+094D  U+0928 |
| द्ध  U+0926 U+094D U+0927 | द्घ  U+0926 U+094D U+0918 | |
| ष्ट  U+0937 U+094D U+091F | ष्ठ  U+0937 U+094D U+0920 | |
| द्व  U+0926 U+094D U+ 0935 | द्ब  U+0926 U+094D U+092C | |

We also noted that the display of the composite characters varies across fonts. While the same may appear similar and confusing in one font, in the other font they might appear quite distinct. Below, we provide some examples. The confusingly similar composite characters in the respective fonts are highlighted in yellow:

खोक्रो खोक्नो (In Kalimati)

द्व द्ब (Aerial Unicode MS)

द्व द्ब (Mangal)

द्व द्ब (Kalimati)

As per our observation, the confusability of composite characters is more prominent in Aerial Unicode MS font.

We also found out that the vowel character “ई” could be typed in two different ways in Unicode – the first one being typing it the usual way using the Unicode keyboard mapping (U+0908) and the second one being typing it using the following combination of characters (र + ् + इ = र्इ (U+0930 + U+094D + U+0907). Apparently, this way of typing has its root in the traditional ASCII based fonts, where the “ई” used to be produced by first typing “इ” and then attaching a “reph” symbol above. Our recommendation is that we disallow the production of the “reph” symbol by using the following combinations (र + ् + any independent vowel) as the actual designation of the “reph” symbol is to denote a short “ra” form as in the words कर्म, धर्म etc.

# Homophonic variants

## Words involving the ciravindu:

According to the writing practice in Nepali, the siravindu character is allowed just in front of the characters - य, र, ल, व, श, स, ष, ह only but this rule does not necessarily apply for other languages in the Devanagari script.

For example, in Nepali:

सम्बन्धी, हिन्दी, ढुङ्गा, शङ्का (Preferred way of writing)

संबन्धी, हिंदी, शंका, ढुंगा, डुंगा, टंकण (Incorrect way of writing)

In general while writing in Nepali, the “cirabindu” and the combination (म + ्) are interchangeably used. For example, संवत and सम्वत. Hence, we would recommend the words formed by the following character combinations to be kept as a “blocked” variant:

Character + ं **+ [**य, र, ल, व, श, स, ष, ह], for example, संवत

Character + म **+** ् **+ [**य, र, ल, व, श, स, ष, ह], for example, सम्वत

However, the following combination of characters is incorrect:

Character + ं **+** म **+** ्**+** character, for example, संम्वत

## Words involving halanta:

In Nepali, there is growing trend to use or drop the “halanta” character at the end of word and vice-versa. There are just a limited set of such words in the Nepali dictionary.

For example,

परिषद् vs परिषद

संवत् vs संवत

सांसद् vs सांसद

Any word with/without halanta ending should be treated as a “blocked” variant.

## Words involving the visarga:

Any word with/without the “visarga” character should be treated as a “blocked” variant. We provide a few examples below:

दुःख

दुख

अतः

अत

प्रायशः

प्रायश

निःशुल्क

निशुल्क