

Devanagari i-kar(dependent vowel 'i') rending problem - a simple solution.

- Hariram Pansari <hariraama@gmail.com>

Richard Ishida <ishida@w3.org> posted at:

<<http://www.w3.org/International/tests/repository/css3-selectors/first-letter/results-first-letter#devanagari>>

Devanagari text

Snapshot summary, 2014-02-14

[Firefox 26.0, Chrome 32.0.1700.107, Safari 6.1.1, Opera 19.0, Internet Explorer 9]

Only Firefox selects what would equate to a grapheme cluster, rather than a single character at the start of a line.

None of the browsers support selection of conjoined consonants. These are basic units in indic scripts, and all component characters need to be highlighted together, but this goes beyond the definition of a grapheme cluster in the Unicode Standard.

These results mean that `::first-letter` is of little use for languages such as Hindi and Marathi at the moment, even on Firefox. Text in those languages does use things such as drop caps at the start of a paragraph, but that would therefore have to be coded into the HTML manually using spans.

NOTE: Firefox can sometimes appear to support more complex sequences. This doesn't happen in these tests, which download a WOFF font to show the devanagari characters. But if, on a Mac, the Devanagari MT system font is applied to the test (see [this version of the test](#)), the following correct selections are produced.

स्ति उड प्रियबर

This appears to be very font specific: it works with the Devanagari MT font, but not with other fonts on the system used. It also only works when a left-sighted vowel-sign is used. The other conjuncts on the page failed.

It also appears to be Firefox specific.

Assertion	Firefox	Chrome	Safari	Opera	IE	Detailed results
hindi, letter only	Pass	Fail	Fail	Fail	Fail	css3-selectors-first-letter-010
hindi, base character with 1 combining character	Pass	Fail	Fail	Fail	Fail	css3-selectors-first-letter-011
hindi, consonant + 2 combining characters	Pass	Fail	Fail	Fail	Fail	css3-selectors-first-letter-012
hindi, simple conjunct	Fail	Fail	Fail	Fail	Fail	css3-selectors-first-letter-014
hindi, conjunct plus combining character	Fail	Fail	Fail	Fail	Fail	css3-selectors-first-letter-016

Devanagari dependent vowel 'i' (ि) is traditionally placed at left side of the base consonant, but logically it is inputted, stored and processed after the base consonants. (As all matras=dependent vowels are stored and processed after the base consonants). So re-ording and re-positioning of the glyph is required and complex algorithms at OS-level (Like Uniscribe in Windows and Pango in Linux) are used for achieving this.

In Windows OS if usp10.dll is absent or de-activated, the ि (U+093F) will render after the base word.

The word

किस्मत

will render as

किस्मत

Which will not be accepted by general public, as tragically ि should appear left side of the character.

This problem of complex rendering cannot be solved at Open Type font level only. Because all the possible syllables(अक्षर or पूर्णाक्षर) of Devanagari with ि matra has be designed and stored in the OT font itself and those can be rendered by glyph replacement technique. But creating all the glyphs of all such possible syllables (including possible conjuncts) is almost impossible task. So this has to be made with special algorithm in the Unicode Script Processing Engine of the OS.

A simpler solution:

The complex rendering problem of the Devanagari ि मात्रा matra can be solved with this simpler approach.

This matra is to be divided in two parts.

The left site appearing bar (कण or दण्ड) + the top cap placed after the base consonants:

ि = ि + ि = ि ि

ति = ि + त + ि

The 'danda' placed before the consonant can be stored/denoted/represented with Unicode Char U+094E (should appear without the dotted circle U+25CC)

094E - DEVANAGARI VOWEL SIGN PRISHTHAMATRA E

- character has historic use only
- combines with E to form AI, with AA to form O, and with O to form AU

The top cap placed after the consonant can be stored with Unicode Char U+093F (as it is used now in all OT fonts). ि

The technique of typing after and appearing above/below/on the previous character is now already in use in 8bit TTF fonts and 16 bit OT fonts by keeping the glyph left side of the outline of the Character space in Fontographer or Fontforge. (As in case of matra like ँ ॠ ॡ ॢ ॣ । (0941, 0942, 0943, 0947, 0948, 0901, 0902) etc.

Sorting and Indexing:

The left side bar denoted with U+094E can be ignored in sorting easily and the top cap only denoted for ि matraa only should be used in sorting and indexing.

Benefits:

This will make Devanagari script processing simple. If the USP10.DLL or PANGO or likewise other OS rendering engines are absent, the Devanagari ि matra will render correctly above the previous consonants.

खिगि....हि

ि matra on conjunct syllables:

Indian Standard DEVANAGARI SCRIPT AND HINDI SPELLINGS IS 16500:2012 clause No.4.1.2.5 states:

"हल् चिह्न युक्त वर्ण से बननेवाले संयुक्ताक्षर के द्वितीय व्यंजन के साथ 'इ' की मात्रा का प्रयोग संबंधित व्यंजन के तत्काल पूर्व ही किया जाए, न कि पूरे युग्म के पूर्व; जैसे- कुट्टिम, चिट्ठियाँ, द्वितीय, बुद्धिमान, चिह्नित (कुट्टिम, चिट्ठियाँ, द्वितीय, बुद्धिमान, चिह्नित नहीं।)"

So representation of ि matra over a consonant with hal-sign (viram U+094D) is wrong. (Scientifically and logically wrong). As a consonant with hal-sign means that it does not have any vowel inherent. So how can the ि matra will be put on a pure consonant? So correct form is कुट्टिम, चिट्ठियाँ, द्वितीय, बुद्धिमान, चिह्नित etc.

Again as per the definition in

Indian Languages layout requirements(Hindi as initial language)

<<http://w3cindia.in/Indic-req-draft/Revised-Indic-layout-requirements.html>>

ABNF definition of Indic syllable

Rule 2 : {CH}C[v][m]

Sl. No.	Examples	Definition
1.	र, क, ज, ल, म	Consonant is a syllable
2.	प्प,कख,क्त, ज्ज्व, त्कल, त्स्न	Zero or more Consonant + Virama sequences followed by consonant is a syllable
3.	र्त, र्त्स, र्त्स्न, र्त्स्न्या, फ्रक	Zero or more Consonant (Nukta) +Virama followed by consonant is a syllable
4.	र्ता, र्त्स्न्या, फ्रजी, क्या	Zero or more consonant+ (Nukta)+ <u>virāma</u> sequences followed by a

A Vowel is always placed on a syllable at last (after the vowel, only vowel-modifier can be placed). In a syllable a dependent-vowel(matra) cannot come in between two pure consonants (or Consonants with hal-sign or viram).

So the use of matra of इ = ि

seems logically and scientifically wrong in the words like:

क्कि, क्खि, स्सि, ज्जि, प्पि, त्थि

Placing of ि matra should be on the last consonant only is right sequence.

क्कि, क्खि, ससि, जजि, पपि, तथि

But this will not be accepted by general public, because

-- the lower portion i.e. danda or bar of the ि matra will appear as 'कण kana' i.e. the joining bar of a letter, removing which half-conjunct (= +्) or halant-yukta or viram-yukta consonants are formed.

क् = क

ख् = ख

ग् = ग

ह् = ह

So any matra(dependent vowel) cannot be put on any half-consonant.

But, IS 16500:2012 has no any clause regarding clarifying this issue. This is very important question. Devanagari become complex by this unscientific rendering. **This Indian Standard of BIS, needs thorough revision.**

But we have to make a Balance between traditional uses of the script and technically and scientifically correct processing of the script on computers. Therefore complex rendering engine at OS-level has made.

My suggestion to form ि matra with two part glyphs as above (094E+093F-top-cap) may simplify the Devanagari layout.

(Solve of reph-reordering problem follows...)

- Hariram Pansari <hariraama@gmail.com>