

### Feedback on Draft Devanagari Script Behaviour for Hindi Version 1.4.10

S. No.	Feedback/ Remark From TDIL-DC Portal Users	Comments
1.	<p>The definition of Indic syllable has been revised as under :</p> <p>V[m]  {CH}C[v][m] CH</p> <p>The Linguistic definition of Indic syllable has been mapped to ABNF(Augmented Backus–Naur Form) for the purpose of text segmentation, Line breaking , Drop letter, letter spacing in horizontal text and vertical text representation. The definition has been elaborated taking Hindi as an example.</p> <p><b>The definition is combination of 3 rules :</b></p> <p>Rule 1 : V[m]</p> <p>Rule 2 : {CH}C[v][m]</p> <p>Rule 3 : CH (This rule is applicable only at the end of the word)</p> <p>V(Upper case) is complete vowel m is modifier(Anusvara/Visarga/Chandrabindu) C is Consonant as per Unicode definition which may or may not include nukta v (lower case) is any dependent vowel or vowel sign (mātrā) H is halant / virama   is a rule seperator [ ] - The enclosed items is optional under this bracket { } - The enclosed item/items occurs once or repeated multiple times</p> <p>Examples:</p> <p>Rule 1 : V[m]</p> <p>Sl. No.</p>	

	<p>Examples</p> <p>Definition</p> <p>1.</p> <p>अ, ई, उ</p> <p>V (Vowel) is a syllable</p> <p>2.</p> <p>अं, ऊँ, आः</p> <p>V+ Modifier is a syllable</p> <p>Rule 2 : {CH}C[v][m]</p> <p>Sl. No.</p> <p>Examples</p> <p>Definition</p> <p>1.</p> <p>र, क, ज, ल, म</p> <p>Consonant is a syllable</p> <p>2.</p>	
--	---	--

प्प,कख,क्त, ज्ज्व, त्क्ल,त्स्न

Zero or more Consonant + Virama sequences followed by consonant is a syllable

3.

र्त्त, त्स, त्सर्न्, त्सर्न्त्य, फ्रक्

Zero or more Consonant (Nukta) +Virama followed by consonant is a syllable

4.

र्त्ता, त्सर्न्त्या, फ्रजी, क्या

Zero or more consonant+ (Nukta)+ virāma sequences followed by a consonant  
(+Nukta) followed by a vowel sign is a syllable

5.

तः,स्तं, स्त्रँ, स्तः, फ्रजँ

zero or more consonant+ (Nukta)+ virāma sequences followed by a consonant  
(+Nukta) followed by modifier is a syllable

6.

त्सर्न्त्याः त्सर्न्त्युं, त्सर्न्त्युँ, फ्रजँ,हिं

zero or more consonant+ (Nukta)+ virāma sequences followed by a consonant (+Nukta) followed by a vowel sign and modifier is a syllable

7.

स्थि,ज्जि,ख्वा

Zero or more Consonant +halant sequences followed by a consonant followed by vowel sign is a syllable

Rule 3 : CH

त्, व्, म्, भ् etc are syllable in Hindi only at the end of the word

Examples of combination of the rules :

1. स्वागतम् - CHCv + C + C + CH has following syllables :

स्वा

CHCv

ग

C

त

C

म्

CH

2. भरतनाट्यम्-  $C + C + C + C_v + CHC + C$

भ

C

र

C

त

C

ना

$C_v$

ट्य

CHC

म

C

	<p>3. सद्बुद्धि - C + CHC<sub>v</sub> + CHC<sub>v</sub></p> <p>स</p> <p>C</p> <p>बु</p> <p>CHC<sub>v</sub></p> <p>द्धि</p> <p>CHC<sub>v</sub></p> <p>The proposed definition is generic in nature and has already being tested for 11 Indian languages i.e Hindi, Marathi, Bengali, Nepali, Tamil, Telugu, Kannada, Gujarati, Punjabi, Oriya &amp; Malayalam. The new rule for CH(Consonant+ Halant) occurrence at the end of the word has been introduced. The testing of the remaining languages is underway.</p> <p>From  <b>Prashant Verma I Sr. Software Engineer</b>  <b>W3C India</b></p>	
2.	<p>Please refer Annexure-1 for for suggestion on Draft Standard Devanagari Script Behavior.</p> <p>From  <b>Mahesh Chander Vashisth</b></p>	

3.	<p>1. Insofar as Akshar is concerned, once it is frozen, we can modify the text accordingly.</p> <p>2. Also the current Indic definition included is having Devanagri examples only and they should be language specific. (Clarification sought on the above statement: which section is being referred to ?)</p> <p>3. Section for developers was possible for Hindi because CHD had the Manak Hindi document for Hindi. Such documents do not exist for other languages and hence the section for developers has been left out. Request TDIL, DeitY to help us to locate such documents, if available from the respective states. As far as our knowledge goes most of them do not have it.</p> <p>From <b>Mahesh Kulkarni</b> <b>Associate Director and HoD, GIST</b></p>	
4.	<p>CDAC has provided script behaviour documents (Part A) for six languages + Hindi and others are in the pipe-line. However as mentioned in our earlier mail, unlike Hindi, Part B (Guidelines for Developers) cannot be provided for the other languages , since to the best of our knowledge State IT's do not have documentation pertaining to the same.</p> <p>In a discussion with you on 25th July 2014, the following was decided.</p> <p>1. Part B comprises of two sections:</p> <p>1. Technical Guidelines</p> <p>2. Linguistic Information pertaining to the language</p>	

	<p>1. Insofar as the technical guidelines were concerned, it was decided as under:</p> <p>SECTION</p> <p>HEAD</p> <p>RESPONSIBILITY/CONTACT INSTITUTION</p> <p>1.1 – 1.2</p> <p>Script and Historical approach</p> <p>Community</p> <p>State IT Secretary</p> <p>1.3 TO 1.6.</p> <p>Encoding Principles and</p> <p>Akshar</p> <p>Information about Akshar is provided in each Language in Part A in Section 6.2.</p> <p>It is recommended that the same be removed from Part B in the case of Hindi.</p> <p>TDIL to provide the same</p> <p>1.7</p> <p>UAX Segmentation Rules</p>	
--	---	--



	<p>TDIL to provide the same</p> <p>1.8</p> <p>Rendering rules</p> <p>Detailed description not available for other languages in Ch. 9.0 of Unicode. IT secretaries be requested to provide the same.</p> <p>1.9</p> <p>ZWJ/ZWNJ</p> <p>Made available in Annexure 2 for each language.</p> <p>1.10</p> <p>Cursor Movement and Deletion</p> <p>These are derived from Akshar and it was felt that Microsoft be contacted to provide the same for other languages. The Hindi rules were taken from the Microsoft Site</p> <p>1.11</p> <p>Normalization</p> <p>This section was specifically provided for Hindi where this affects the script.</p> <p>No such rules exist for other languages. However Unicode Normalisation rules be</p>	
--	---	--

referred to. <http://unicode.org/reports/tr15/>

1. Insofar as Linguistic Information is confirmed, it is requested that each State IT secretary be requested to provide the same, as mandated by the respective state government for that particular language. It was also suggested that the Style Guides prepared by CIIL be referred to by the State Government for mandating the said linguistic information.

It was felt that the following disclaimer be put on the Website for the Script Behaviour documents:

#### DISCLAIMER

The script behaviour document comprises a set of recommendations laid down by the experts for the use of the community. It does not purport in any way to be prescriptive in nature nor is it to be interpreted as a Standard.

In case of any difference of opinion, please provide feedback on the portal.  
Your contribution will be appreciated.

From

**Mahesh Kulkarni**

**Associate Director and HoD, GIST**

## Annexure-1

देवनागरी लिपि व्यवहार 1.4.10 के संदर्भ में कुछ सुझाव इस प्रकार हैं:-

कां.			मात्रा के ये प्रकार नहीं होते हैं।		
गलत रूप		सही रूप			
अ		आ			
इ		ई			
उ		ऊ			
ए		ऐ			
ओ		औ			
प्र		प्र			
प		प्र			

  

क्रम सं.	व्यंजन के संयुक्त होने के प्रकार	उदाहरण	क्रम सं.	व्यंजन के संयुक्त होने के प्रकार	उदाहरण
1	ह + ः + व	जिह्वा	10	त + ः + त	उत्तर
2	ह + ः + ल	प्रह्लाद	11	त + ः + त + ः	तत्त्व
3	ह + ः + न	चिह्न	12	क + ः + त	भक्त
4	ह + ः + म	ब्रह्म	13	ङ + ः + क	अङ्क
5	ह + ः + ण	ह्ण	14	ङ + ः + ख	शङ्ख
6	ह + ः + य	बाह्य	15	ङ + ः + ग	अङ्ग
7	द + ः + र	श्रीमद्भगवद्गीता	16	ङ + ः + घ	कङ्घा
8	ह + ः + र	ह्रस्व	17	ञ + ः + च	कञ्चा
9	द + ः + भ	अद्भुत	18	ञ + ः + ज	पञ्जा