

# **SCRIPT GRAMMAR FOR HINDI LANGUAGE**

**Prepared by**

**Technology Development for Indian Languages (TDIL) Programme**

**Department of Information Technology, Government of India**

**in association with**

**Centre for Development of Advanced Computing (C-DAC)**

## Table of Contents

0. INTRODUCTION .....	3
1. OBJECTIVES OF SCRIPT GRAMMAR .....	4
2. END USERS FOR SCRIPT GRAMMAR .....	5
3. SCOPE .....	6
4. TERMINOLOGY .....	7
5. PHILOSOPHY AND UNDERLYING PRINCIPLES.....	11
6. SCRIPT GRAMMAR STRUCTURE.....	12
6.1. PERIPHERAL ELEMENTS OF THE SCRIPT GRAMMAR.....	13
6.2. CONFORMITY TO THE SYLLABLE STRUCTURE .....	14
6.3 SCRIPT GRAMMAR PROPER.....	17
6.3.1. The Character Set of Hindi. ....	17
6.3.2. Consonant Mātrā Combinations. ....	24
6.3.3. The Ligature Set of Hindi. ....	38
6.3.4 The Collation Order of Hindi.....	48
7. REFERENCES .....	50
8. ANNEXURES .....	51
Annexure 1: Names of experts who have contributed to the script grammar.....	51
Annexure 2: Shapes of Hindi characters as per Central Hindi Directorate .....	52
Annexure 3: Unicode Table of Devanāgarī .....	54

## 0. INTRODUCTION

The term **script grammar** refers to the behavior pattern of the writing system of a given language. Languages which have written representations do not use a haphazard manner of storing the information within the system, but use a coherent pattern which is similar to the linguistic grammar of a given language. With the help of specialists (not necessarily linguist) who work in the area of the written representation of the language, the manner in which the shapes of the characters of the language and the representation of the conjunct forms is provided. In other words, the Script Grammar deals with the surface structure of the language and tries to provide the best possible “fit” for shapes and their representation. Since this is a highly subjective issue, the shapes provided here are recommendations at the best and conform to the perception of the mandating body/evaluators who consensually arrive at the “best possible fit” which is acceptable to a majority of users. An example from the Devanāgarī script will make the above clear. Although Marathi and Hindi share the same script Devanāgarī, not only do they not share the same character inventory but in addition the representation of certain characters is different. Thus the Hindi /la/ is different from the Marathi /la/ in so far as the placement of the stem is concerned Hindi ल Marathi ल. This ensures that the Script Grammar conforms to the language in question and provides the character shapes acceptable to a given user community. It should be noted that this does not mean monotony. The Marathi and Hindi /la/ can have a variety of forms once the intrinsic structure of the character is determined.

Script Grammar is the term used to define:

- the writing system used to inscribe a given language
- the syllabic structure of the writing system of the language
- the rule ordering of the characters within the akshar
- description of the syllabic clusters / ligatures
- collation order of the characters: lexical / dictionary sorting order

## 1. OBJECTIVES OF SCRIPT GRAMMAR

The objectives of the script grammar for each language can be divided into two major parts:

### **Societal:**

- Provide a visual representation of shapes that are deemed to be in conformity with the perception of a given community
- Ensure thereby that this perception is safe-guarded
- Through wide-spread dissemination and creation of appropriate tools ensure that within the given linguistic community, all media try to adopt the given shape.

### **Technical:**

- Classify the language in terms of its ISO and also whether it belongs to the Abjad, Akshar (Alphasyllabary) class.
- Provide an inventory of the characters pertinent to the language and classify the same in terms of their taxonomy.
- As a corollary determine whether the inventory is in conformity to the Syllable formalism as stipulated in ISCII'91 and subsequently adopted by Unicode.
- Since Brahmi is written from left to right, and since certain characters do not follow the linear L to R order, provide an inventory of displaced concatenators, i.e. characters such as Mātrās that concatenate to the Consonant
- Propose the best shape representation of the individual characters as well as of the ligatures used within a given script. As a corollary request the expert(s) to identify the largest possible strings of such ligatures.
- Finally provide the collation order pertinent to that script/language, which would be of great utility to high-end NLP as well as to CLDR's in the pertinent language. The collation order for Hindi is different from Marathi although both languages share the same script. Thus, in Hindi क्ष, ज्ञ are placed at the end of the consonant inventory, i.e. after ह in the sort order. In Hindi क्ष is sorted along with क and ज्ञ with ज

## 2. END USERS FOR SCRIPT GRAMMAR

The script grammar specific to a given language can be used by a large number of users.

- The primary aim of the script grammar has been for the font developer. The grammar can be used by font developers desirous of developing a font which is compliant with the perception of the characters and ligatures of a language by its user community.
- Certain features of the script grammar such as the shapes can also be used for testing OCR and OHWR. Similarly information regarding ligatures as well as collation order can help in high-end NLP work such as detecting invalid combinations, correct implementation of syllable structure, prediction routines to name a few. Information regarding collation and character sets can be also used for CLDR.
- They allow the font designer to design a font which is in compliance with the norms and standards of that particular script. A major problem which will be dealt with in the template is one of ligatures. The final list of ligatures defined by the script grammar allows the font designer to write specific rules for such glyphs.
- It permits the software developer to design and implement the keyboard and the input mechanism which will meet the requirement of the particular linguistic community.
- The collation or sort order as described in a Script Grammar permits the software developer to write software functions/ routines for sorting data in all applications.
- Script grammars are equally important for keyboard design, especially when supplemented by frequency data from a corpus.

As can be seen the script grammar has a wide range of use and can be of utility to font developers, Indian language developers and linguists in the area of computation.

### 3. SCOPE

This script grammar document contains following information about the language and the script used for writing the language.

1. Name of the language and its representation in the 3 letter mnemonic as per ISO 639.1 & 639.3 standard.
2. Script used to inscribe the given language
3. The structure of the script used for writing the language
  - Rule ordering of the characters within the syllable formation is a language
  - Description of the syllabic clusters of the script
  - Collation order of the characters: lexical / dictionary sorting order
  - Compliance of the script with Unicode.

These will be treated within the relevant sections of the script grammar.

## 4. TERMINOLOGY<sup>1</sup>

**Abjad:** A writing system in which each symbol always or usually stands for a consonant. The long vowels are indicated. However the short vowels are rarely marked and the reader needs to supply these. Example: Urdu written in Perso-Arabic Script is an example of this writing system.

**Abugida:** Also called an alphasyllabary, it is a segmental writing system in which consonant–vowel sequences are written as a unit: each unit is based on a consonant letter, and vowel notation is obligatory, but secondary.<sup>2</sup>

**Akshar:** see **Abugida**.

**Allographs:** Variants of the representation of a character. Thus ae and æ [U+00E6] in Latin alphabet are allographs.

**AlloScript:** The term relates to languages which share a common script. Thus Devanāgarī is used to write 9 official languages. However these languages do not use the same set of characters. Thus Marathi uses the retroflex lla ऌ [U+ 0933] which Hindi does not use. Flaps used in Hindi ञ [U+095C] ण [U+095D] are not used in Konkani. These sub-sets of scripts based on a single “matricial” script are termed as alloscripts.

**Alphabet:** A set of letters used in writing a language. Example: The English alphabet.

**Aspirated consonant:** A consonant which is pronounced with an extra puff of air coming out at the time of release of the oral obstruction. This has a sound of an extra "h".

**Basic alphabet:** The minimal set of letters which can be used for uniquely encoding every word of a language. The basic alphabet for English consists of only the upper-case letters A-Z.

**Catenators:** Also termed as concatenators, these are characters which are concatenated to another character. In the Brahmi script these are the Mātrās or Vowel modifiers which are adjoined to the consonant and add a vocalic value to the consonant.

**Conjunct:** The Brahmi derived scripts are noted for a large number of consonant conjunct forms that serve as orthographic abbreviations (ligatures) of two or more adjacent letterforms. This abbreviation takes place only in the context of a consonant cluster. Under normal circumstances, a consonant cluster is depicted with a conjunct glyph if such a glyph is available in the current font. In the absence of a conjunct glyph,

---

<sup>1</sup> As in the case of the BIS Document, in order to make the terminology accessible for all readers, examples have been chosen from English/Latin scripts, wherever possible. Some definitions have been excerpted from the BIS ISCI91 document and suitably modified where necessary.

<sup>2</sup> Wikipedia definition

the one or more dead consonants that form part of the cluster are depicted using half-form glyphs. In the absence of half-form glyphs, the dead consonants are depicted using the nominal consonant forms combined with visible virama signs.<sup>3</sup>

**Consonant:** A letter representing a speech sound in which the flow of air is at least partly obstructed in the oral tract.

**Diacritic:** A mark added to a letter which distinguishes it from the same letter without a mark, usually having a different phonetic value or stress.

**Displaced Catenator:** (see Catenator) Within the Brahmi script, the writing system is linear and moves from left to right. However, in the case of some catenators this rule is not observed and the catenator (wholly or partially) is placed to the right of the consonant to which it relates. The short vowel I / ॠ / in Devanāgarī is an example of a displaced catenator.

**Display composing:** It is the process of organizing the basic shapes available in a font in order to display (or print) a word.

**Display rendition:** It is the process by which a string of characters is displayed (or printed). In this process several consecutive characters may combine with each other on the screen. The sequence of display of the characters may become different.

**Eyebrow repha:** (See **Eyesh ra**).

**Eyesh ra:** The eyesh ra is used in Konkani, Nepali and Marathi. It is treated as different from the ॠ (repha) by certain linguists. While the former is treated as a flap, the latter is a continuant trill (*cf.*, Kalyan Kale and Anjali Soman 1986).

**Font:** A set of symbols used for display or printing of a script in a particular style.

**International numerals:** The conventional 0 to 9 digits used in English for denoting numbers. These are also known as Indo-Arabic numerals (to differentiate them from the Roman numerals like IX for 9).

**Latin alphabet:** The alphabet used for writing the language of ancient Rome. Also known as the Roman alphabet. The alphabet is used today for writing English and European languages and also many Indian languages..

**Letter:** A character representing one or more of the simple or compound sounds used in speech. It can be any of the alphabetic symbols.

**Ligature:** (see **Conjunct**).

---

<sup>3</sup> Unicode ver. 6.0 Chapter 9.0 pp. 6-7.



**Nasal consonant:** A consonant pronounced with the flow of air passing through the nose and the mouth. Example *m, n* in English.

**Nasalized vowel:** A vowel pronounced with the flow of air passing both through the nose and the mouth. In Indian scripts this is denoted by a Chandrabindu and gives the vowel/vowel modifier over which it is placed a nasalised value. Example: जाँच .

**Phonetic alphabet:** An alphabet which has direct correspondence between letters and sounds Example: The International Phonetic Alphabet.

**Pure consonant:** A consonant which does not have any vowel implicitly associated with it.

**Rafar:** A special case of a ligature constituted by the adjunction of ra followed by a halanta to consonant. The resultant combination places the ra on top of the consonant to which it is adjoined. In case the consonant itself is adjoined to another consonant, the rafar is placed above the consonant e.g. र्+क कँ , र्+घ्+य घ्यँ.

**Rakar:** A special case of a ligature constituted by the adjunction of a consonant followed by a halanta to ra. In a large number of Brahmi derived scripts the ra is adjoined to the stem of consonant to which it relates. In the case of consonants which have no stem such as the retroflexes in Devanāgarī, the rakar is placed below the consonant to which it relates.

**Repha:** (see **Rafar**).

**Roman script:** The script based on the ancient Roman alphabet, with the letters A-Z and additional diacritic marks. Used for writing a language which is not usually written in the Roman alphabet.

**Script:** A distinctive and complete set of characters used for the written form of one or more languages.

**Script numerals:** The 0 to 9 digits in a script, which have shapes distinct from their international counterparts.

**Syllable:** A unit of pronunciation uttered without interruption, forming whole or part of a word, and usually having one vowel or diphthong sound optionally surrounded by one or more consonants.

**Transliteration:** Representation of words with the closest corresponding letters in an alphabet of a different language.

**Vowel:** A letter representing a speech sound made with the vibration of the vocal cords, but without audible obstruction. In some languages voiceless vowels do occur.

**Vowel sign/allograph:** A graphic character associated with a letter, to Brahmi derived  
ate a vowel to be associated with that character (Mātrā in Hindi).

## 5. PHILOSOPHY AND UNDERLYING PRINCIPLES

The script grammar is based on the following principles:

1. The Grammar aims to depict the surface grammar of the written language: the manner in which characters as well as conjuncts are depicted.
2. Where a given script admits many languages, it is pre-supposed that such languages will prescribe different representations for a given shape or conjunct according to the perception of the native users of that language.
3. Corollary to the above, the result is a script and alloscripts i.e. a given script shared by many languages is not uniformly deployed across all the languages, but is subject to variations and modulations.
4. The term Grammar is used here in a non-normative sense: what is prescribed is in the form of recommendations provided by experts who visualize the shape of the given script in their mother tongue in a specific manner. Subjective variations may occur<sup>4</sup>.
5. The Grammar is limited to its synchronic use, i.e. the manner in which a given language as of today admits a character set within the script used to write it. It is not diachronic or historical in nature and does not study the evolution of the given script across centuries.

---

<sup>4</sup> It is recommended that such variations be culled by placing the Grammars of different scripts in public review.

## **6. SCRIPT GRAMMAR STRUCTURE**

The script grammar provided below has the following parts.

Part 6.1. deals with peripheral elements such as the ISO of the language, the writing system used: (Alphasyllabic) Abugida or Abjad.

Part 6.2. treats of the syllabic structure. It verifies whether the character set of the language complies with the ISCII syllabic structure and if not, which cases are not compliant.

Part 6.3 is the script grammar proper and describes the character set as well as the conjunct shapes of the given script along with the collation order.

## 6.1. PERIPHERAL ELEMENTS OF THE SCRIPT GRAMMAR

These constitute the elements that are peripheral to the script grammar. The main parameters considered are the mnemonic and name of the language (needed for CLDR and also for language tags), the writing system used to inscribe the language and wherever possible a short history of the language.

### 6.1.1. Name of the language and its representation in the 3 letter mnemonic as per ISO 639.1. & 639.3

**Name of the Language: HINDI**

**ISO Mnemonics: *hin***

This refers to a one line description of the language and its mnemonic representation as per the ISO.

### 6.1.2. Identification of the writing system(s) used to inscribe the given language **Hindi is written using the Devanāgarī script. It is an alphasyllabary with the akshar as its core.**

This is a one line description of the script used to write the language. However, in case the language uses more than one script, all the scripts in question are specified, provided these constitute the official language of the given state.

All scripts derived from Brahmi are Abugidas, i.e. syllabary driven systems. The main features of Abugidas are as under:

- The consonant has an implicit vowel built-in which is normally the schwa.
- The inherent vowel can be modified by the addition of other vowels or muted by a diacritic termed as a Virama or Halanta.
- Vowels can be handled as full vowels with a vocalic value .
- When two or more consonants join together they form ligatures which can be recognized by their shape कल or alternatively form an entirely new shape क्ष = क्ष.

Abugidas/Alphasyllabaries because of their syllabic structure require a special description which is the subject of the discussion in 6.2. below.

### 6.1.3. Amendments needed in Unicode for Hindi language

None has been proposed by the experts who have mandated the script grammar.

## 6.2. CONFORMITY TO THE SYLLABLE STRUCTURE

**Hindi language complies with the akshar structure described above. It can admit up to 3 consonant clusters.**

Alphasyllabaries are determined by the notion of the Akshar. The compositional grammar of the syllable determines its well-formedness. This is through a series of formal constraints based on a Backus-Naur Formalism which is given below. The akshar, first defined in the ISCII document (1991), identifies the following character ‘sub-sets’ for the purposes of identifying the akshar. In what follows the syllable analysis will be restricted to Hindi.

### (C) Consonants

क	क्	ख	ख	ग
ग	घ	ङ	च	छ
ज	झ	झ	ञ	ट
ठ	ड	ड	ढ	ढ
ण	त	थ	द	ध
न	प	फ	फ़	ब
भ	म	य	र	ल
व	श	ष	स	ह

### (V) Vowels

अ	आ	इ	ई	उ	ऊ	ऋ	ए	ऐ	ऐ	ओ	औ	औ
---	---	---	---	---	---	---	---	---	---	---	---	---

### (M) Mātrās or Vowel Modifiers

ा	ि	ी	ु	ू	ृ	े	ँ	ै	ो	ॉ	ौ
---	---	---	---	---	---	---	---	---	---	---	---

### (D) Diacritics

ं: Anuswar	Anuswara, an archinasal, is denoted by a dot above the letter after which it is to be pronounced. This falls under Nasal category.
ँ: Chandrabindu	Chandrabindu, a nasal, is denoted by a breve with a dot superposed above the letter after which it is to be pronounced. This falls under Nasal category.

◌ः Visarga	Visarga, denoted by two dots placed one above the other.
S: Avagraha <sup>5</sup>	For extra length with long vowels as seen in the Sanskrit text /उपदेशेऽजनुनासिक इत् /

(H):Halanta ◌् - Halanta is used in most writing systems to signify the lack of an inherent vowel.

(N)<sup>6</sup> Nukta ◌̣ - Nukta is used in Hindi

Each of these sub-types has its restrictions in terms of what can precede or follow it, within an akshar, as shown in the table below:

PRECEDED BY	SUBTYPE	FOLLOWED BY
-, H	<b>C</b>	N,M,D,H
C	<b>N</b>	C,M,D,H
-	<b>V</b>	D
C, N	<b>M</b>	D
C, N,V,M	<b>D</b>	-
C, N	<b>H</b>	C

C can be preceded by H or no subtype and followed by any one of the following: M,D,H

N can be preceded by C and followed by any one of the following: N,M,D,H

V can be preceded by no subtype and followed by D but not by another sub-type.

M can be preceded by C and followed by D.

D can be preceded by C, V, M and followed by no other subtype. It closes the akshar.

H can be preceded by C alone and followed only by C and no other sub-set.

### 6.2.1.Akshar Types

The formalism defines the akshar in terms of both what can constitute an akshar and what cannot. A valid akshar as per this definition can be of only two types:

1. A vowel akshar : a full vowel.
2. A consonant akshar : a full consonant (having a mātrā )

The four other subsets viz. Mātrās, Vowel Modifiers, Halanta and Nukta cannot constitute an akshar by themselves or in combination among themselves.

1. The Vowel akshar is of the following types:

- 1.1. A pure vowel all by itself: अ, /a/ आ /ā/ etc.

<sup>5</sup> Avagraha is rarely used in Hindi and is used for representing text in Sanskrit here the character is needed..

<sup>6</sup> The nukta is a small dot placed under a character in certain scripts to show that they are flapped or for deriving consonants required for Urdu क़,ख़,ग़,ज़,फ़

1.2. A vowel followed by a modifier, i.e. either an archinasal (*anuswara*) or a visarga: ई /ī/ , आ: /āH/

2. The Consonant akshar can be of the following types:

2.1. A full consonant (with or without Nukta), i.e. with the inherent vowel : क : /ka/

2.2. A consonant<sup>7</sup> (with or without Nukta) followed by a mātrā, i.e. the inherent vowel being substituted by another vowel: की /ki:/, की /qi:/

2.3. A consonant (with or without Nukta) followed by a modifier: कं /k̄/, हः /h̄aH<sup>8</sup>/

2.4. A consonant (with or without Nukta) followed by a mātrā and a modifier: कुं /kū/, दुः /duH/.

2.5. A consonant cluster i.e. a half consonant (Consonant+Halanta) followed by a full consonant followed optionally by a mātrā, a modifier or a combination of both. These result in a ligature or what is often termed as *yuktakshara*.

त्क /tka/ त्कं /tk̄/, त्कः /tkaH/ त्कुं /tkū/, त्दु /tdu/.

The above permutations and combinations result in 7 major akshar types. Of these the last type introduces the problem of the number of consonant clusters. ISCII (91, p.23) provides for up to three consonant clusters as the “worst case”, i.e. the largest possible string. This is functional for Modern Prakrits where the largest consonantal cluster rarely exceeds three consonant. Sanskrit is an exception where in a single word, five consonants can come together: कात्स्न्य /kārtsnya/ "wholeness", "entirety" (secondary derivative from the adjective कृत्स्न /kṛtsna/ meaning “whole, complete”).)

This means that theoretically the following forms can be postulated:

1. Vowel Set: With the Vowel as the node.

V VD

2. Consonant set: With the Consonant as the node (an implicit or modified vowel is pre-implied).

Node	Mātrā	Modifier	Mātrā+Modifier
C <sup>9</sup>	CM	CD	CMD
CHC	CHCM	CHCD	CHCMD
CHCHC	CHCHCM	CHCHCD	CHCHCMD
CHCHCHC	CHCHCHCM	CHCHCHCD	CHCHCHCMD

A total number of 16 theoretical syllables is therefore possible.

Since the formal structure script grammar of the akshar is common to all Brahmi based scripts, it will not be treated in the sample template, but it will form the basis of an exhaustive description of the characters as well as their ligatural representations.

<sup>7</sup> For purposes of Simplification, C here will automatically be treated as being also consonant+nukta C+ N

<sup>8</sup> This character represents phonetically the weak implicit vowel, termed as schwa and often shown as /a/ also.

<sup>9</sup> C here will automatically be treated as being also consonant+nukta, C+N to simplify the explanation



## 6.3 SCRIPT GRAMMAR PROPER

This section lays down in detail the different parameters of the Script Grammar for Hindi. These are:

- 6.3.1. The Character Set of Hindi.
- 6.3.2. The Consonant mātrā combinations of Hindi.
- 6.3.3. The Ligature Set of Hindi.
- 6.3.4. Collocation Order of Hindi

### 6.3.1. The Character Set of Hindi.

This section provides detailed information about the characters in the language and the list of the same and also more importantly shows the manner in which the character is to be written. Each subsection comprises therefore two parts: the basic character set and the shape each character should have, as mandated by the experts who have designed the script grammar of Hindi.

This comprises the following:

- 6.3.1.1. The Consonant Set.<sup>10</sup>
- 6.3.1.2. The Vowel Set.<sup>9</sup>
- 6.3.1.3. The Mātrā Set.<sup>9</sup>
- 6.3.1.4. Displaced Catenators.
- 6.3.1.5. Shape of the combination of ra (rakar, repha).
- 6.3.1.6. The Set of Diacritics.
- 6.3.1.7. Numerals.
- 6.3.1.8. Punctuation marks.
- 6.3.1.9. Other symbols.

Each of these will be analysed in detail:

#### 6.3.1.1. The Consonant Set

The Consonant set of Hindi comprises the following characters:

A basic Consonant inventory arranged as per their Vargas.

	STOPS				NASAL	FRICATIVE	
	-voiced -aspirated	-voiced +aspirated	+voiced -aspirated	+voiced +aspirated	+voiced -aspirated	-voiced	+voiced
Laryngeal						ह	
Pharyngeal							ख
Uvular	क <sup>11</sup>		ग				

<sup>10</sup> The shapes provided here are as desired by Central Hindi Directorate. These are provided in Appendix 1

<sup>11</sup> क ग are Not recognized by the Central Hindi Directorate cf. p.2 of their document

Velar	क	ख	ग	घ	ङ		
Palatal	च	छ	ज	झ	ञ		ञ
Retroflex	ट	ठ	ड	ढ	ण		ष
Flaps			ड	ढ			
Dental	त	थ	द	ध	न	स	श
Labio-Dental						फ	
Bi-labial	प	फ	ब	भ	म		

Other consonants

य	र	ल	व
---	---	---	---

**Note:** Ligatures क्ष त्र ज्ञ श्र are not listed in the consonants list<sup>12</sup>.

The exact shapes as desired by the experts are provided in the table below:

	STOPS				NASAL	FRICATIVE	
	-voiced -aspirated	-voiced +aspirated	+voiced -aspirated	+voiced +aspirated	+voiced -aspirated	-voiced	+voiced
Laryngeal						ह	
Pharyngeal							ख
Uvular	क्र <sup>13</sup>		ग				
Velar	क	ख	ग	घ	ङ		
Palatal	च	छ	ज	झ	ञ		ञ
Retroflex	ट	ठ	ड	ढ	ण		ष
Flaps			ड	ढ			
Dental	त	थ	द	ध	न	स	श
Labio-Dental						फ	

<sup>12</sup> These are treated as Ligatures by the Central Hindi Directorate cf. p 2 and also p..5 of their document

<sup>13</sup> क्र ग are Not recognized by the Central Hindi Directorate cf. p.2 of their document

Bi-labial	प	फ	ब	भ	म		
-----------	---	---	---	---	---	--	--

Other consonants

य	र	ल	व
---	---	---	---

### 6.3.1.2. The Vowel Set

The Vowel set of Hindi is as under:

अ	DEVANAGARI LETTER A
आ	DEVANAGARI LETTER AA
इ	DEVANAGARI LETTER I
ई	DEVANAGARI LETTER II
उ	DEVANAGARI LETTER U
ऊ	DEVANAGARI LETTER UU
ऋ	DEVANAGARI LETTER VOCALIC R
ए	DEVANAGARI LETTER E
ऐ	DEVANAGARI LETTER AI
ओ	DEVANAGARI LETTER O
औ	DEVANAGARI LETTER AU
ऐँ	DEVANAGARI LETTER CHANDRA E
औँ	DEVANAGARI LETTER CHANDRA O

As per expert recommendations the character set should be written as under:

अ आ इ ई उ ऊ ऋ ए ऐ ऐँ ओ औँ

### 6.3.1.3. The Mātrā Set

The Mātrā<sup>14</sup> (Vowel Sign) of Hindi is as under:

Mātrā Names	Mātrās Sign	Where is it used ?	Consonant Shapes formed
1. Devanāgarī sign AA	ा	आ	क् + आ = का
2. Devanāgarī sign I ( stands to the left of the consonant)	ि	इ	क् + इ = कि
3. Devanāgarī sign II	ी	ई	क् + ई = की
4. Devanāgarī sign U	ु	उ	क् + उ = कु
5. Devanāgarī sign UU	ू	ऊ	क् + ऊ = कू
6. Devanāgarī sign vocalic R	ृ	ऋ	क् + ऋ = कृ
7. Devanāgarī sign E	े	ए	क् + ए = के
8. Devanāgarī sign AI	ै	ऐ	क् + ऐ = कै
9..Devanāgarī sign O	ो	ओ	क् + ओ = को
10. Devanāgarī sign AU	ौ	औ	क् + औ = कौ
11 Devanāgarī sign Candra E	ँ	ऐँ	क् +ँ = कैँ
12 Devanāgarī sign Candra O	ॉ	औँ	क् +ॉ = कौँ

As per the experts' recommendations the character set should be written as under:

ा ि ी उ ू ृ े ँ ै ो ौ  
ॉ

---

<sup>14</sup> Allographs

#### 6.3.1.4. Displaced Catenators

Under normal circumstances Vowel Modifiers also known as catenators (since they concatenate to the preceding consonant) in Brahmi based scripts are written from left to right in linear order (with the exception of Consonant stacks). However, certain modifiers are displaced and are placed to the left of the consonant to which they concatenate. As a general rule in all Devanāgarī script driven languages there is only one displaced catenator:

CATENATOR	POSITION	EXAMPLE
ि	To l e f t o f char act er	कि , रि , पि

#### 6.3.1.5. Shape of the combination of ra (rakar, repha)

The र takes a variety of shapes known as rakar and repha (rafar) depending on its position. When conjoined before a consonant by means of the halanta, it changes shape and is placed on top of the consonant or consonant clusters to which it relates. This is called a repha or rafar. When it is conjoined after a consonant with the help of a halanta, it appends to the consonant in the shape of a slanting stroke attached to the stem (side rakar) or in the case of consonants which have no stem such as ट, it is appended in the shape of a ^ to the bottom of the character (bottom rakar). Hindi has the following combinations of ra:

##### RAFARS

Top rafar: के ते e.g. top rafars will be formed in case of following words.

धर्म , चर्खा
--------------

##### RAKARS

1. Bottom rakar      ट      ड
2. Side rakar        क्र      च

Examples of words where Rakars are used in Hindi are given below:

Bottom rakar	ड्रम , राष्ट्र
Side rakar	व्रत , चक्र , प्रसाद , स्रोत

#### 6.3.1.6. Diacritics

These are as under in the case of Hindi:

ं : - Anuswar रंग

ँ : - Candr abi ndu/ Anunasi ka ईँट

ः - Hal ant संवत्

ः - V sarga दुःख

s : - For extra length with long vowel s e.g. / उपदेशेऽजनुनासिक इत्/ <sup>15</sup>

### 6.3.1.7. Numerals

The international number set (Latin-Arabic set : 0,1,2,3,4,5,6,7,8,9) is used in official documents in Hindi.

Following are the numbers used in Hindi.

० १ २ ३ ४ ५ ६ ७ ८ ९

Numeral Shapes	Explanation
०	Devanāgarī Digit Zero
१	Devanāgarī Digit One
२	Devanāgarī Digit Two
३	Devanāgarī Digit Three
४	Devanāgarī Digit Four
५	Devanāgarī Digit Five
६	Devanāgarī Digit Six
७	Devanāgarī Digit Seven
८	Devanāgarī Digit Eight
९	Devanāgarī Digit Nine

### 6.3.1.8. Punctuation Markers

Hindi uses punctuation markers from the Latin set. such as . , ; : “ ‘ ( ) [ ] etc.

However, the abbreviation marker is often used in Devanagari .

Purna and Deerga Virama (full-stop/danda) Devanagari code block: U+0964, U+0965

।,॥ । is used to mark the full stop and । and ॥ are used for poetry as in the Doha of Sant Kabir.

<sup>15</sup> Avagraha is rarely used in Hindi and is used for representing text in Sanskrit here the character is needed..

चाह मिटी, चिंता मिटी मनवा बेपरवाह । जिसको कुछ नहीं चाहिए वह शहनशाह॥  
माटी कहे कुम्हार से, तु क्या रौंदे मोय । एक दिन ऐसा आएगा, मैं रौंदूंगी तोय ॥

A list of punctuations is provided below:

Sr. No.	Name of the marker	Marker Shape
1.	Full Stop or Period	.
2.	Question Mark	?
3.	Comma	,
4.	Exclamation Mark	!
5.	Apostrophe	'
6.	Semi Colon	;
7.	Colon	:
8.	Hyphen	-
9.	Dash	--
10.	Ellipsis mark	...
11.	Oblique	/
12.	Double quotation mark	" "
13.	Single quotation mark	‘ ’
14.	Cross	XXX
15.	As Above	-- " --
16.	Round Brackets	( )
17.	Square Brackets	[ ]
18.	Curly Brackets	{ }
19.	Abbreviation Sign	(.)
20.	Devanāgarī Danda	
21.	Devanāgarī Double Danda	

### 6.3.1.9 Other Symbols

These are religious, currency markers etc. included in Unicode:

ॐ: Om (as written in Hindi)

₹: Rupee Sign as mandated by Government of India.

### 6.3.2. Consonant Mātrā Combinations.

These refer to the shapes generated when a Mātrā is adjoined to the Consonant. The layout of these is in the shape of a matrix where the first horizontal row refers to the active consonant and the first vertical column refers to the vowel-modifier.

Due to constraints of space and also for reasons of clarity, for each class a series of 3 tables are provided.

Table 1: क ख ग घ ङ च छ ज झ ञ



Table 2: ट ठ ड ढ ण त थ द ध न

Table 3: प फ ब भ म य र ल व श

ष स ह

Table 4: क ख ग ज ङ इ ऋ

Wherever there is an X it implies that the combination does not exist. For the font developer this is an indication that for this particular combination which is not possible in the language but needs to be accommodated in the font table, a simple linear combination be provided. The script grammar document provides therefore for such combinations which could be termed as “dead” but which are required by the font developer for developing the font for the particular language.

e.g. Although the combination of ङ + Mātrā is theoretically not possible it needs to be handled at the font level in the anticipation that a user could type this combination. The font would show the following: ङ

The classes are as under:

6.3.2.1. refers to a simple concatenation of Consonant and Mātrā combinations.

6.3.2.2. refers to a concatenation of Consonant and Mātrā + Nasal marker combinations.

These are with Anuswara and Chandrabindu. Other diacritics such as avagraha and visarga have been avoided, since these are linear in nature, are adjoined to the combination and do not in any way modify the structure of the shapes.

### 6.3.2.1 Consonant and Mātrā combinations.

This set refers to a simple concatenation of Consonant and Mātrā.

#### Consonant and Mātrā combinations Set 1

	क	ख	ग	घ	ङ	च	छ	ज	झ	ञ
ा	का	खा	गा	घा	×	चा	छा	जा	झा	×
ि	कि	खि	गि	घि	×	चि	छि	जि	झि	×
ी	की	खी	गी	घी	×	ची	छी	जी	झी	×
ु	कु	खु	गु	घु	×	चु	छु	जु	झु	×
ू	कू	खू	गू	घू	×	चू	छू	जू	झू	×
ृ	कृ	खृ	गृ	घृ	×	चृ	छृ	जृ	झृ	×
े	के	खे	गे	घे	×	चे	छे	जे	झे	×
ै	कै	खै	गै	घै	×	चै	छै	जै	झै	×
ो	को	खो	गो	घो	×	चो	छो	जो	झो	×
ॉ	कॉ	खॉ	गॉ	घॉ	×	चॉ	छॉ	जॉ	झॉ	×
ौ	कौ	खौ	गौ	घौ	×	चौ	छौ	जौ	झौ	×

Remark 1- ङ and ञ are rarely used only as the first members of clusters and mostly as परसवर्ण or alternatives of अनुस्वार

## Consonant and Mātrā combinations Set 2

This set is in continuation of set 1 which shows consonant and Matra combinations.

	ट	ठ	ड	ढ	ण	त	थ	द	ध	न
ा	टा	ठा	डा	ढा	x	ता	था	दा	धा	ना
ि	टि	ठि	डि	ढि	x	ति	थि	दि	धि	नि
ी	टी	ठी	डी	ढी	x	ती	थी	दी	धी	नी
ु	टु	ठु	डु	ढु	x	तु	थु	दु	धु	नु
ू	ट्र	ठ्र	ड्र	ढ्र	x	त्र	थ्र	द्र	ध्र	न्र
ृ	टृ	ठृ	डृ	x	x	x	थृ	दृ	धृ	नृ
े	टे	ठे	डे	ढे	x	ते	थे	दे	धे	ने
ै	टै	ठै	डै	ढै	x	तै	थै	दै	धै	नै
ो	टो	ठो	डो	ढो	x	तो	थो	दो	धो	नो
ॉ	टॉ	ठॉ	डॉ	ढॉ	x	तॉ	थॉ	दॉ	धॉ	नॉ
ौ	टौ	ठौ	डौ	ढौ	x	तौ	थौ	दौ	धौ	नौ

### Consonant and Mātrā combinations Set 3

This set is in continuation of set 2 which shows consonant and Matra combinations.

	प	फ	ब	भ	म	य	र	ल	व	श	ष	स	ह
ा	पा	फा	बा	भा	मा	या	रा	ला	वा	शा	षा	सा	हा
ि	पि	फि	बि	भि	मि	यि	रि	लि	वि	शि	षि	सि	हि
ी	पी	फी	बी	भी	मी	यी	री	ली	वी	शी	षी	सी	ही
ु	पु	फु	बु	भु	मु	यु	रु	लु	वु	शु	षु	सु	हु
ू	पू	फू	बू	भू	मू	यू	रू	लू	वू	शू	षू	सू	हू
ृ	पृ	फृ	बृ	भृ	मृ	X	X	X	वृ	शृ	X	सृ	हृ
े	पे	फे	बे	भे	मे	ये	रे	ले	वे	शे	षे	से	हे
ै	पै	फै	बै	भै	मै	यै	रै	लै	वै	शै	षै	सै	है
ो	पो	फो	बो	भो	मो	यो	रो	लो	वो	शो	षो	सो	हो
ॉ	पॉ	फॉ	बॉ	भॉ	मॉ	यॉ	रॉ	लॉ	वॉ	शॉ	षॉ	सॉ	हॉ

#### Consonant and Mātrā combinations Set 4

This set is in continuation of set 3 which shows consonant and Matra combinations for Nukta Consonants.<sup>16</sup>

	क	ख	ग	ज	फ	ड	ढ
ा	का	खा	गा	जा	फा	डा	ढा
ि	कि	खि	गि	जि	फि	डि	ढि
ी	की	खी	गी	जी	फी	डी	ढी
ु	कु	खु	गु	जु	फु	डु	ढु
ू	कू	खू	गू	जू	फू	डू	ढू
ृ <sup>17</sup>	कृ	खृ	गृ	जृ	फृ	डृ	ढृ
े	के	खे	गे	जे	फे	डे	ढे
ै	कै	खै	गै	जै	फै	डै	ढै
ो	को	खो	गो	जो	फो	डो	ढो
ॉ	कॉ	खॉ	गॉ	जॉ	फॉ	डॉ	ढॉ
ौ	कौ	खौ	गौ	जौ	फौ	डौ	ढौ

---

<sup>16</sup> क,ग are not recognized as part of the Consonant set by the Central Hindi Directorate.

<sup>17</sup> This combination is rarely used but still noted as a guide for the font designer.

### 6.3.2.2 Consonant and Mātrā +Nasal combinations.

This set refers to a Consonant and Mātrā + Nasal marker combinations.

#### Consonant and Mātrā + Nasal combinations - Set 1

	क	ख	ग	घ	ङ	च	छ	ज	झ	ञ
ं	कं	खं	गं	घं	ङं	चं	छं	जं	झं	ञं
ां	कां	खां	गां	घां	ङां	चां	छां	जां	झां	जां
िं	किं	खिं	गिं	घिं	ङिं	चिं	छिं	जिं	झिं	जिं
ीं	कीं	खीं	गीं	घीं	ङीं	चीं	छीं	जीं	झीं	जीं
ुं	कुं	खुं	गुं	घुं	ङुं	चुं	छुं	जुं	झुं	जुं
ूं	कूं	खूं	गूं	घूं	ङूं	चूं	छूं	जूं	झूं	जूं
ृं	कृं	खृं	गृं	घृं	ङृं	चृं	छृं	जृं	झृं	जृं
ें	कैं	खैं	गैं	घैं	ङैं	चैं	छैं	जैं	झैं	जैं
ँ	कँ	खँ	गँ	घँ	ङँ	चँ	छँ	जँ	झँ	जँ
ैं	कैं	खैं	गैं	घैं	ङैं	चैं	छैं	जैं	झैं	जैं
ों	कों	खों	गों	घों	ङों	चों	छों	जों	झों	जों
ौं	काँ	खाँ	गाँ	घाँ	ङाँ	चाँ	छाँ	जाँ	झाँ	जाँ
ौं	कौं	खौं	गौं	घौं	ङौं	चौं	छौं	जौं	झौं	जौं

### Consonant and Mātrā +Nasal combinations - Set 2

This set is in continuation of set 1 above which shows combinations of Consonant and Mātrā + Nasal marker

	ट	ठ	ड	ढ	ण	त	थ	द	ध	न
ं	टं	ठं	डं	ढं	णं	तं	थं	दं	धं	नं
ां	टां	ठां	डां	ढां	णां	तां	थां	दां	धां	नां
िं	टिं	ठिं	डिं	ढिं	णिं	तिं	थिं	दिं	धिं	निं
ीं	टीं	ठीं	डीं	ढीं	णीं	तीं	थीं	दीं	धीं	नीं
ुं	टुं	ठुं	डुं	ढुं	णुं	तुं	थुं	दुं	धुं	नुं
ूं	ट्रूं	ठ्रूं	ड्रूं	ढ्रूं	ण्रूं	त्रूं	थ्रूं	द्रूं	ध्रूं	न्रूं
ृं	टृं	ठृं	डृं	ढृं	णृं	तृं	थृं	दृं	धृं	नृं
ें	टे	ठे	डे	ढे	णे	ते	थे	दे	धे	ने
ँ	टँ	ठँ	डँ	ढँ	णँ	तँ	थँ	दँ	धँ	नँ
ैं	टै	ठै	डै	ढै	णै	तै	थै	दै	धै	नै
ों	टों	ठों	डों	ढों	णों	तों	थों	दों	धों	नों
ॉं	टाँ	ठाँ	डाँ	ढाँ	णाँ	ताँ	थाँ	दाँ	धाँ	नाँ
ौं	टौं	ठौं	डौं	ढौं	णौं	तौं	थौं	दौं	धौं	नौं

<sup>18</sup> Since ँ,ँ are used only for the purpose of writing loan words mainly from English. However, since such occurrences are rare, the nasal shapes of ँँ,ँँ are rarely encountered in Hindi.

### Consonant and Mātrā +Nasal combinations - Set 3

This set is in continuation of set 2 above which shows combinations of Consonant and Mātrā + Nasal marker

	प	फ	ब	भ	म	य	र	ल	व	श	ष	स	ह
ं	पं	फं	बं	भं	मं	यं	रं	लं	वं	शं	षं	सं	हं
ां	पां	फां	बां	भां	मां	यां	रां	लां	वां	शां	षां	सां	हां
िं	पिं	फिं	बिं	भिं	मिं	यिं	रिं	लिं	विं	शिं	षिं	सिं	हिं
ीं	पीं	फीं	बीं	भीं	मीं	यीं	रीं	लीं	वीं	शीं	षीं	सीं	हीं
ुं	पुं	फुं	बुं	भुं	मुं	युं	रुं	लुं	वुं	शुं	षुं	सुं	हुं
ूँ	पूँ	फूँ	बूँ	भूँ	मूँ	यूँ	रूँ	लूँ	वूँ	शूँ	षूँ	सूँ	हूँ
ृँ	पृँ	फृँ	बृँ	भृँ	मृँ	X	X	X	वृँ	शृँ	षृँ	सृँ	हृँ
ेँ	पेँ	फेँ	बेँ	भेँ	मेँ	येँ	रेँ	लेँ	वेँ	शेँ	षेँ	सेँ	हेँ
ँ <sup>19</sup>	पँ	फँ	बँ	भँ	मँ	यँ	रँ	लँ	वँ	शँ	षँ	सँ	हँ
ैँ	पैँ	फैँ	बैँ	भैँ	मैँ	यैँ	रैँ	लैँ	वैँ	शैँ	षैँ	सैँ	हैँ
ोँ	पोँ	फोँ	बोँ	भोँ	मोँ	योँ	रोँ	लोँ	वोँ	शोँ	षोँ	सोँ	होँ
ौँ	पोँ	फौँ	बौँ	भौँ	मौँ	योँ	रौँ	लौँ	वौँ	शौँ	षौँ	सौँ	हौँ
ौँ	पोँ	फौँ	बौँ	भौँ	मौँ	योँ	रौँ	लौँ	वौँ	शौँ	षौँ	सौँ	हौँ

<sup>19</sup> Since ँ,ँ are used only for the purpose of writing loan words mainly from English. However, since such occurrences are rare, the nasal shapes of ँँ,ँँ are rarely encountered in Hindi.



### Consonant and Mātrā +Nasal combinations - Set 4

This set is in continuation of set 3 above which shows combinations of Consonant and Mātrā + Nasal marker

	क	ख	ग	ज	फ	ड	ढ
ं	कं	खं	गं	जं	फं	डं	ढं
ा	का	खा	गा	जा	फा	डा	ढा
िं	किं	खिं	गिं	जिं	फिं	डिं	ढिं
ीं	कीं	खीं	गीं	जीं	फीं	डीं	ढीं
ुं	कुं	खुं	गुं	जुं	फुं	डुं	ढुं
ूं	कूं	खूं	गूं	जूं	फूं	डूं	ढूं
ृं	कृं	खृं	गृं	जृं	फृं	डृं	ढृं
ें	कें	खें	गें	जें	फें	डें	ढें
ँ <sup>20</sup>	कँ	खँ	गँ	जँ	फँ	डँ	ढँ
ैं	कैं	खैं	गैं	जैं	फैं	डैं	ढैं
ों	कों	खों	गों	जों	फों	डों	ढो
ॉं	काँ	खाँ	गाँ	जाँ	फाँ	डाँ	ढाँ
ौं	कौं	खौं	गौं	जौं	फौं	डौं	ढौ

<sup>20</sup> Since ँ, ॉ are used only for the purpose of writing loan words mainly from English. However, since such occurrences are rare, the nasal shapes of ँ, ॉ are rarely encountered in Hindi.

## Consonant and Mātrā + Nasal combinations: With Chandrabindu<sup>21</sup> - Set 1

	क	ख	ग	घ	ङ	च	छ	ज	झ	ञ
ँ	कँ	खँ	गँ	घँ	ङँ	चँ	छँ	जँ	झँ	ञँ
ाँ	काँ	खाँ	गाँ	घाँ	ङाँ	चाँ	छाँ	जाँ	झाँ	जाँ
िँ	किँ	खिँ	गिँ	घिँ	ङिँ	चिँ	छिँ	जिँ	झिँ	ञिँ
ीँ	कीँ	खीँ	गीँ	घीँ	ङीँ	चीँ	छीँ	जीँ	झीँ	जीँ
ुँ	कुँ	खुँ	गुँ	घुँ	ङुँ	चुँ	छुँ	जुँ	झुँ	जुँ
ूँ	कूँ	खूँ	गूँ	घूँ	ङूँ	चूँ	छूँ	जूँ	झूँ	जूँ
ृँ	कृँ	खृँ	गृँ	घृँ	ङृँ	चृँ	छृँ	जृँ	झृँ	जृँ
ेँ	केँ	खेँ	गेँ	घेँ	ङेँ	चेँ	छेँ	जेँ	झेँ	जेँ
ैँ	कैँ	खैँ	गैँ	घैँ	ङैँ	चैँ	छैँ	जैँ	झैँ	जैँ
ोँ	कोँ	खोँ	गोँ	घोँ	ङोँ	चोँ	छोँ	जोँ	झोँ	जोँ
ौँ	काँ	खाँ	गाँ	घाँ	ङाँ	चाँ	छाँ	जाँ	झाँ	जाँ
ौँ	कोँ	खोँ	गोँ	घोँ	ङोँ	चोँ	छोँ	जोँ	झोँ	जोँ

<sup>21</sup> As per rule 2.6.2.3. of the Central Hindi Directorate, Chandrabindu cannot be placed over matras which are above the Shirorekha and in this case the Chandrabindu is replaced by an Anuswar.

### Consonant and Mātrā +Nasal combinations With Chandrabindu - Set 2

This set is in continuation of set 1 above which shows combinations of Consonant and Mātrā + Chandrabindu.

	ट	ठ	ड	ढ	ण	त	थ	द	ध	न
ँ	टँ	ठँ	डँ	ढँ	णँ	तँ	थँ	दँ	धँ	नँ
ाँ	टाँ	ठाँ	डाँ	ढाँ	णाँ	ताँ	थाँ	दाँ	धाँ	नाँ
िँ	टिँ	ठिँ	डिँ	ढिँ	णिँ	तिँ	थिँ	दिँ	धिँ	निँ
ीँ	टीँ	ठीँ	डीँ	ढीँ	णीँ	तीँ	थीँ	दीँ	धीँ	नीँ
ुँ	टुँ	ठुँ	डुँ	ढुँ	णुँ	तुँ	थुँ	दुँ	धुँ	नुँ
ूँ	ट्रूँ	ठ्रूँ	ड्रूँ	ढ्रूँ	ण्रूँ	त्रूँ	थ्रूँ	द्रूँ	ध्रूँ	न्रूँ
ृँ	ट्रृँ	ठ्रृँ	ड्रृँ	ढ्रृँ	ण्रृँ	त्रृँ	थ्रृँ	द्रृँ	ध्रृँ	न्रृँ
ेँ	टेँ	ठेँ	डेँ	ढेँ	णेँ	तेँ	थेँ	देँ	धेँ	नेँ
ैँ	टैँ	ठैँ	डैँ	ढैँ	णैँ	तैँ	थैँ	दैँ	धैँ	नैँ
ोँ <sup>22</sup>	टोँ	ठोँ	डोँ	ढोँ	णोँ	तोँ	थोँ	दोँ	धोँ	नोँ
ॉँ	टाँ	ठाँ	डाँ	ढाँ	णाँ	ताँ	थाँ	दाँ	धाँ	नाँ
ौँ	टौँ	ठौँ	डौँ	ढौँ	णौँ	तौँ	थौँ	दौँ	धौँ	नौँ

<sup>22</sup> Since ँ, ॉ are used only for the purpose of writing loan words mainly from English. However, since such occurrences are rare, the nasal shapes of ैँ, ौँ are rarely encountered in Hindi.

**Consonant and Mātrā +Nasal combinations With Chandrabindu - Set 3**

This set is in continuation of set 2 above which shows combinations of Consonant and Mātrā + Chandrabindu.

	प	फ	ब	भ	म	य	र	ल	व	श	ष	स	ह
ँ	पँ	फँ	बँ	भँ	मँ	यँ	रँ	लँ	वँ	शँ	षँ	सँ	हँ
ाँ	पाँ	फाँ	बाँ	भाँ	माँ	याँ	राँ	लाँ	वाँ	शाँ	षाँ	साँ	हाँ
िँ	पिँ	फिँ	बिँ	भिँ	मिँ	यिँ	रिँ	लिँ	विँ	शिँ	षिँ	सिँ	हिँ
ीँ	पीँ	फीँ	बीँ	भीँ	मीँ	यीँ	रीँ	लीँ	वीँ	शीँ	षीँ	सीँ	हीँ
ुँ	पुँ	फुँ	बुँ	भुँ	मुँ	युँ	रुँ	लुँ	वुँ	शुँ	षुँ	सुँ	हुँ
ूँ	पूँ	फूँ	बूँ	भूँ	मूँ	यूँ	रूँ	लूँ	वूँ	शूँ	षूँ	सूँ	हूँ
ृँ	पृँ	फृँ	बृँ	भृँ	मृँ	यृँ	रृँ	लृँ	वृँ	शृँ	षृँ	सृँ	हृँ
ेँ	पेँ	फेँ	बेँ	भेँ	मेँ	येँ	रेँ	लेँ	वेँ	शेँ	षेँ	सेँ	हेँ
ैँ	पेँ	फेँ	बेँ	भेँ	मेँ	येँ	रेँ	लेँ	वेँ	शेँ	षेँ	सेँ	हेँ
ौँ	पोँ	फोँ	बोँ	भोँ	मोँ	योँ	रोँ	लोँ	वोँ	शोँ	षोँ	सोँ	होँ
ॉँ	पाँ	फाँ	बाँ	भाँ	माँ	याँ	राँ	लाँ	वाँ	शाँ	षाँ	साँ	हाँ
ौँ	पोँ	फोँ	बोँ	भोँ	मोँ	योँ	रोँ	लोँ	वोँ	शोँ	षोँ	सोँ	होँ

**Consonant and Mātrā +Nasal combinations With Chandrabindu - Set 4**

This set is in continuation of set 3 above which shows combinations of Consonant and Mātrā + Chandrabindu.

	क	ख	ग	ज	फ	ड	ढ
ँ	कँ	खँ	गँ	जँ	फँ	डँ	ढँ
ाँ	काँ	खाँ	गाँ	जाँ	फाँ	डाँ	ढाँ
िँ	किँ	खिँ	गिँ	जिँ	फिँ	डिँ	ढिँ
ीँ	कीँ	खीँ	गीँ	जीँ	फीँ	डीँ	ढीँ
ुँ	कुँ	खुँ	गुँ	जुँ	फुँ	डुँ	ढुँ
ूँ	कूँ	खूँ	गूँ	जूँ	फूँ	डूँ	ढूँ
ृँ	कृँ	खृँ	गृँ	जूँ	फृँ	डृँ	ढृँ
ेँ	केँ	खेँ	गेँ	जेँ	फेँ	डेँ	ढेँ
ैँ	कैँ	खैँ	गैँ	जैँ	फैँ	डैँ	ढैँ
ोँ	कोँ	खोँ	गोँ	जोँ	फोँ	डोँ	ढोँ
ॉँ	काँ	खाँ	गाँ	जाँ	फाँ	डाँ	ढाँ
ौँ	कोँ	खोँ	गोँ	जोँ	फोँ	डोँ	ढोँ

### 6.3.3. The Ligature Set of Hindi.

Hindi has a large set of ligatural forms. These are combinations of Consonant+Halanta+Consonant (CHC) or CHCHC or even rarer CHCHCHC. The CHC combinations which are the most frequent are arranged in the shape of a matrix: the abscissa or horizontal axis refers to the Consonant which constitutes the ligature and the ordinate or vertical axis shows the consonant which forms the ligature and which is followed by a halanta.

As in 6.3.2. the ligature sets are divided into the following

6.3.3.1 CHC (in a matrix)

6.3.3.2 CHCHC

6.3.3.3.CHCHCHC

#### 6.3.3.1. CHC ( combination of two Consonants)

These ligatures are presented as in the earlier case of Consonant+Mātrā combinations in three sets. A lot of “slots” have an X marked, showing that the ligature is not possible in the language but is theoretically possible. In these cases, the font developer is to assume that the ligature is linear in nature.

The following set shows a combination of two consonants. To know how particular combinations forms, select one consonant from the first column and second from first row. For e.g. Combination of consonant “क” and “क” is ligature “क्क”.

CHC( combination of two consonants) - Set 1										
	क	ख	ग	घ	ङ	च	छ	ज	झ	ञ
क्	क्क	क्ख	क्ग	क्घ	क्ङ	क्च	क्छ	क्ज	क्झ	क्ञ
ख्	ख्क	ख्ख	ख्ग	ख्घ	ख्ङ	ख्च	ख्छ	ख्ज	ख्झ	ख्ञ
ग्	ग्क	ग्ख	ग्ग	ग्घ	ग्ङ	ग्च	ग्छ	ग्ज	ग्झ	ग्ञ
घ्	घ्क	घ्ख	घ्ग	घ्घ	घ्ङ	घ्च	घ्छ	घ्ज	घ्झ	घ्ञ
ङ्	ङ्क	ङ्ख	ङ्ग	ङ्घ	ङ्ङ	ङ्च	ङ्छ	ङ्ज	ङ्झ	ङ्ञ
च्	च्क	च्ख	च्ग	च्घ	च्ङ	च्च	च्छ	च्ज	च्झ	च्ञ
छ्	X	X	X	X	X	X	X	X	X	X
ज्	ज्क	ज्ख	ज्ग	ज्घ	ज्ङ	ज्च	ज्छ	ज्ज	ज्झ	ज्ञ
झ्	झ्क	झ्ख	झ्ग	झ्घ	झ्ङ	झ्च	झ्छ	झ्ज	झ्झ	झ्ञ
ञ्	ञ्क	ञ्ख	ञ्ग	ञ्घ	ञ्ङ	ञ्च	ञ्छ	ञ्ज	ञ्झ	ञ्ञ
ट्	ट्क	ट्ख	ट्ग	ट्घ	ट्ङ	ट्च	ट्छ	ट्ज	ट्झ	ट्ञ
ठ्	ठ्क	ठ्ख	ठ्ग	ठ्घ	ठ्ङ	ठ्च	ठ्छ	ठ्ज	ठ्झ	ठ्ञ
ड्	ड्क	ड्ख	ड्ग	ड्घ	ड्ङ	ड्च	ड्छ	ड्ज	ड्झ	ड्ञ
ढ्	ढ्क	ढ्ख	ढ्ग	ढ्घ	ढ्ङ	ढ्च	ढ्छ	ढ्ज	ढ्झ	ढ्ञ

ण्	ण्क	ण्ख	ण्ग	ण्घ	ण्ड	ण्च	ण्छ	ण्ज	ण्ञ	णञ
त्	त्क	त्ख	त्ग	त्घ	ट्ड	त्च	त्छ	त्ज	त्ज्ञ	त्ञ
थ्	थ्क	थ्ख	थ्ग	थ्घ	थ्ड	थ्च	थ्छ	थ्ज	थ्ञ	थञ
द्	द्क	द्ख	द्ग	द्घ	ट्ड	द्च	द्छ	द्ज	द्ज्ञ	द्ञ
ध्	ध्क	ध्ख	ध्ग	ध्घ	ध्ड	ध्च	ध्छ	ध्ज	ध्ञ	धञ
न्	न्क	न्ख	न्ग	न्घ	न्ड	न्च	न्छ	न्ज	न्ज्ञ	न्ञ
प्	प्क	प्ख	प्ग	प्घ	प्ड	प्च	प्छ	प्ज	प्ज्ञ	प्ञ
फ्	फ्क	फ्ख	फ्ग	फ्घ	फ्ड	फ्च	फ्छ	फ्ज	फ्ञ	फञ
ब्	ब्क	ब्ख	ब्ग	ब्घ	ब्ड	ब्च	ब्छ	ब्ज	ब्ज्ञ	ब्ञ
भ्	भ्क	भ्ख	भ्ग	भ्घ	भ्ड	भ्च	भ्छ	भ्ज	भ्ञ	भञ
म्	म्क	म्ख	म्ग	म्घ	म्ड	म्च	म्छ	म्ज	म्ञ	मञ
य्	य्क	य्ख	य्ग	य्घ	य्ड	य्च	य्छ	य्ज	य्ञ	यञ
र्	र्क	र्ख	र्ग	र्घ	र्ड	र्च	र्छ	र्ज	र्ज्ञ	र्ञ
ल्	ल्क	ल्ख	ल्ग	ल्घ	ल्ड	ल्च	ल्छ	ल्ज	ल्ञ	लञ
व्	व्क	व्ख	व्ग	व्घ	व्ड	व्च	व्छ	व्ज	व्ञ	वञ
श्	श्क	श्ख	श्ग	श्घ	श्ड	श्च	श्छ	श्ज	श्ज्ञ	श्ञ
ष्	ष्क	ष्ख	ष्ग	ष्घ	ष्ड	ष्च	ष्छ	ष्ज	ष्ञ	षञ
स्	स्क	स्ख	स्ग	स्घ	स्ड	स्च	स्छ	स्ज	स्ञ	सञ
ह्	ह्क	ह्ख	ह्ग	ह्घ	ह्ड	ह्च	ह्छ	ह्ज	ह्ञ	हञ
क्	क्क	क्ख	क्ग	क्घ	क्ड	क्च	क्छ	क्ज	क्ज्ञ	क्ञ
ख्	ख्क	ख्ख	ख्ग	ख्घ	ख्ड	ख्च	ख्छ	ख्ज	ख्ञ	खञ
ग्	ग्क	ग्ख	ग्ग	ग्घ	ग्ड	ग्च	ग्छ	ग्ज	ग्ञ	गञ
ज्	ज्क	ज्ख	ज्ग	ज्घ	ज्ड	ज्च	ज्छ	ज्ज	ज्ञ	जञ
फ्	फ्क	फ्ख	फ्ग	फ्घ	फ्ड	फ्च	फ्छ	फ्ज	फ्ञ	फञ
ङ्	ङ्क	ङ्ख	ङ्ग	ङ्घ	ङ्ड	ङ्च	ङ्छ	ङ्ज	ङ्ञ	ङञ
ढ्	ढ्क	ढ्ख	ढ्ग	ढ्घ	ढ्ड	ढ्च	ढ्छ	ढ्ज	ढ्ञ	ढञ

### CHC Set 2:

The following set shows a combination of two consonants. To know how particular combinations forms, select one consonant from the first column and second from first row. For e.g. Combination of consonant “क” and “ट” is ligature “कट”.

CHC( combination of two consonants) - Set 2										
	ट	ठ	ड	ढ	ण	त	थ	द	ध	न
क्	कट	कठ	कड	कढ	कण	क्त	कथ	कद	कध	क्न
ख्	खट	खठ	खड	खढ	खण	ख्त	खथ	खद	खध	ख्न
ग्	गट	गठ	गड	गढ	गण	ग्त	गथ	गद	गध	ग्न
घ्	घट	घठ	घड	घढ	घण	घ्त	घथ	घद	घध	घ्न
ङ्	ङट	ङठ	ङड	ङढ	ङण	ङ्त	ङथ	ङद	ङध	ङ्न
च्	चट	चठ	चड	चढ	चण	च्त	चथ	चद	चध	च्न
छ्	छट	छठ	छड	छढ	छण	छ्त	छथ	छद	छध	छ्न
ज्	जट	जठ	जड	जढ	जण	ज्त	जथ	जद	जध	ज्न
झ्	झट	झठ	झड	झढ	झण	झ्त	झथ	झद	झध	झ्न
ञ्	ञट	ञठ	ञड	ञढ	ञण	ञ्त	ञथ	ञद	ञध	ञ्न
ट्	टट	टठ	टड	टढ	टण	ट्त	टथ	टद	टध	ट्न
ठ्	ठट	ठठ	ठड	ठढ	ठण	ठ्त	ठथ	ठद	ठध	ठ्न
ड्	डट	डठ	डड	डढ	डण	ड्त	डथ	डद	डध	ड्न
ढ्	ढट	ढठ	ढड	ढढ	ढण	ढ्त	ढथ	ढद	ढध	ढ्न
ण्	णट	णठ	णड	णढ	णण	ण्त	णथ	णद	णध	ण्न
त्	तट	तठ	तड	तढ	तण	त्त	तथ	तद	तध	त्न
थ्	थट	थठ	थड	थढ	थण	थ्त	थथ	थद	थध	थ्न
द्	दट	दठ	दड	दढ	दण	दत्त	दथ	द्	द्ध	द्र
ध्	धट	धठ	धड	धढ	धण	ध्त	धथ	धद	धध	ध्न
न्	नट	नठ	नड	नढ	नण	न्त	नथ	न्द	नध	न्न
प्	पट	पठ	पड	पढ	पण	प्त	पथ	पद	पध	प्न
फ्	फट	फठ	फड	फढ	फण	फ्त	फथ	फद	फध	फ्न
ब्	बट	बठ	बड	बढ	बण	ब्त	बथ	बद	बध	ब्न
भ्	भट	भठ	भड	भढ	भण	भ्त	भथ	भद	भध	भ्न
म्	मट	मठ	मड	मढ	मण	म्त	मथ	मद	मध	म्न



य्	यट्	य्ठ	यड्	यढ्	य्ण	यत्	य्थ	यद्	य्ध	यन्
र्	र्ट	र्ठ	र्ड	र्ढ	र्ण	र्त	र्थ	र्द	र्ध	र्न
ल्	लट्	लठ्	लड्	लढ्	लण	लत्	लथ	लद्	लध	लन्
व्	वट्	वठ्	वड्	वढ्	व्ण	वत्	वथ	वद्	वध	वन्
श्	शट्	शठ्	शड्	शढ्	श्ण	श्त्	श्थ	श्द	श्ध	श्न
ष्	षट्	षठ्	षड्	षढ्	ष्ण	षत्	षथ	षद्	षध	षन्
स्	सट्	सठ्	सड्	सढ्	स्ण	सत्	स्थ	सद्	सध	सन्
ह्	हट्	हठ्	हड्	हढ्	ह्ण	हत्	हथ	हद्	हध	हन्
क्	कट्	कठ्	कड्	कढ्	क्ण	क्त	क्थ	क्द	क्ध	क्न
ख्	खट्	खठ्	खड्	खढ्	ख्ण	खत्	खथ	खद्	खध	खन्
ग्	गट्	गठ्	गड्	गढ्	ग्ण	गत्	गथ	गद्	गध	गन्
ज्	जट्	जठ्	जड्	जढ्	ज्ण	जत्	जथ	जद्	जध	जन्
फ्	फट्	फठ्	फड्	फढ्	फ्ण	फत्	फथ	फद्	फध	फन्
झ्	झट्	झठ्	झड्	झढ्	झ्ण	झत्	झथ	झद्	झध	झन्
ढ्	ढट्	ढठ्	ढड्	ढढ्	ढ्ण	ढत्	ढथ	ढद्	ढध	ढन्

### CHC SET 3:

The following set shows a combination of two consonants. To know how a particular combinations forms, select one consonant from the first column and second from first row. For e.g. Combination of consonant “क” and “प” is the ligature “क्प”.

CHC( combination of two consonants) - Set 3													
	प	फ	ब	भ	म	य	र	ल	व	श	ष	स	ह
क्	क्प	क्फ	क्ब	क्भ	क्म	क्य	क्र	क्ल	क्व	क्श	क्ष	क्स	क्ह
ख्	खप	खफ	खब	खभ	खम	ख्य	ख्र	खल	खव	खश	खष	खस	खह
ग्	गप	गफ	गब	गभ	गम	ग्य	ग्र	ग्ल	ग्व	गश	गष	गस	गह
घ्	घप	घफ	घब	घभ	घम	घ्य	घ्र	घल	घव	घश	घष	घस	घह
ङ्	ङप	ङफ	ङब	ङभ	ङम	ङ्य	ङ्र	ङल	ङव	ङश	ङष	ङस	ङह
च्	चप	चफ	चब	चभ	चम	च्य	च्र	चल	चव	चश	चष	चस	चह
छ्	छप	छफ	छब	छभ	छम	छ्य	छ्र	छल	छव	छश	छष	छस	छह
ज्	जप	जफ	जब	जभ	जम	ज्य	ज्र	जल	जव	जश	जष	जस	जह
झ्	झप	झफ	झब	झभ	झम	झ्य	झ्र	झल	झव	झश	झष	झस	झह
ञ्	ञप	ञफ	ञब	ञभ	ञम	ञ्य	ञ्र	ञल	ञव	ञश	ञष	ञस	ञह
ट्	टप	टफ	टब	टभ	टम	ट्य	ट्र	टल	टव	टश	टष	टस	टह
ठ्	ठप	ठफ	ठब	ठभ	ठम	ठ्य	ठ्र	ठल	ठव	ठश	ठष	ठस	ठह
ड्	डप	डफ	डब	डभ	डम	ड्य	ड्र	डल	डव	डश	डष	डस	डह
ढ्	ढप	ढफ	ढब	ढभ	ढम	ढ्य	ढ्र	ढल	ढव	ढश	ढष	ढस	ढह
ण्	णप	णफ	णब	णभ	णम	ण्य	ण्र	णल	णव	णश	णष	णस	णह
त्	तप	तफ	तब	तभ	तम	त्य	त्र	तल	तव	तश	तष	तस	तह
थ्	थप	थफ	थब	थभ	थम	थ्य	थ्र	थल	थव	थश	थष	थस	थह
द्	दप	दफ	दब	दभ	दम	द्य	द्र	दल	द्व	दश	दष	दस	दह
ध्	धप	धफ	धब	धभ	धम	ध्य	ध्र	धल	ध्व	धश	धष	धस	धह
न्	नप	नफ	नब	नभ	नम	न्य	न्र	न्ल	नव	नश	नष	नस	नह
प्	पप	पफ	पब	पभ	पम	प्य	प्र	प्ल	प्व	पश	पष	पस	पह
फ्	फप	फफ	फब	फभ	फम	फ्य	फ्र	फल	फ्व	फश	फष	फस	फह
ब्	बप	बफ	बब	बभ	बम	ब्य	ब्र	ब्ल	ब्व	बश	बष	बस	बह
भ्	भप	भफ	भब	भभ	भम	भ्य	भ्र	भ्ल	भव	भश	भष	भस	भह
म्	मप	मफ	मब	मभ	मम	म्य	म्र	म्ल	म्व	मश	मष	मस	मह

य्	य्प	य्फ	य्ब	य्भ	य्म	य्य	य्र	य्ल	य्व	य्श	य्ष	य्स	य्ह
र्	र्प	र्फ	र्ब	र्भ	र्म	र्य	र्र	र्ल	र्व	र्श	र्ष	र्स	र्ह
ल्	ल्प	ल्फ	ल्ब	ल्भ	ल्म	ल्य	ल्र	ल्ल	ल्व	ल्लश	ल्ल्ष	ल्ल्स	ल्ल्ह
व्	व्प	व्फ	व्ब	व्भ	व्म	व्य	व्र	व्ल	व्व	व्श	व्ष	व्स	व्ह
श्	श्प	श्फ	श्ब	श्भ	श्म	श्य	श्च <sup>23</sup>	श्च	श्च	श्श	श्ष	श्स	श्ह
ष्	ष्प	ष्फ	ष्ब	ष्भ	ष्म	ष्य	ष्प्र	ष्ठ	ष्ठ	ष्ठश	ष्ठ्ष	ष्ठ्स	ष्ठ्ह
स्	स्प	स्फ	स्ब	स्भ	स्म	स्य	स्स	स्ल	स्व	स्श	स्ष	स्स	स्ह
ह्	ह्प	ह्फ	ह्ब	ह्भ	ह्म	ह्य	ह्र	ह्ल	ह्व	ह्श	ह्ष	ह्स	ह्ह
क्	क्प	क्फ	क्ब	क्भ	क्म	क्य	क्क	क्ल	क्व	क्श	क्ष	क्स	क्ह
ख्	ख्प	ख्फ	ख्ब	ख्भ	ख्म	ख्य	ख्र	ख्ल	ख्व	ख्श	ख्ष	ख्स	ख्ह
ग्	ग्प	ग्फ	ग्ब	ग्भ	ग्म	ग्य	ग्र	ग्ल	ग्व	गश	ग्ष	ग्स	ग्ह
ज्	ज्प	ज्फ	ज्ब	ज्भ	ज्म	ज्य	ज्र	ज्ल	ज्व	ज्श	ज्ष	ज्स	ज्ह
फ्	फ्प	फ्फ	फ्ब	फ्भ	फ्म	फ्य	फ्र	फल	फ्व	फश	फ्ष	फ्स	फ्ह
ड्	ड्प	ड्फ	ड्ब	ड्भ	ड्म	ड्य	ड्र	ड्ल	ड्व	ड्श	ड्ष	ड्स	ड्ह
ढ्	ढ्प	ढ्फ	ढ्ब	ढ्भ	ढ्म	ढ्य	ढ्र	ढ्ल	ढ्व	ढ्श	ढ्ष	ढ्स	ढ्ह

<sup>23</sup> An important combination very often mis-typed is श+ृ = शृ which is often mis-written as श+र+ ृ शृ. Thus शृंगार is wrongly written as शृंगार.

#### CHC SET 4:

The following set shows a combination of two consonants of which the second form is the nukta . To know how particular combinations form, select one consonant from the first column and second from first row. For e.g. The combination of consonants “क” and “क़” is ligature “क्क़”.

CHC Set 4						
	क	ख	ग	ज	झ	ढ
क	क्क	क्ख	क्ग	क्ज	क्झ	क्ढ
ख	क्क	क्ख	क्ग	क्ज	क्झ	क्ढ
ग	गक	गख	गग	गज	गझ	गढ
घ	घक	घख	घग	घज	घझ	घढ
ङ	ङक	ङख	ङग	ङज	ङझ	ङढ
च	चक	चख	चग	चज	चझ	चढ
छ	छक	छख	छग	छज	छझ	छढ
ज	जक	जख	जग	जज	जझ	जढ
झ	झक	झख	झग	झज	झझ	झढ
ञ	ञक	ञख	ञग	ञज	ञझ	ञढ
ट	टक	टख	टग	टज	टझ	टढ
ठ	ठक	ठख	ठग	ठज	ठझ	ठढ
ड	डक	डख	डग	डज	डझ	डढ
ढ	ढक	ढख	ढग	ढज	ढझ	ढढ
ण	णक	णख	णग	णज	णझ	णढ
त	तक	तख	तग	तज	तझ	तढ
थ	थक	थख	थग	थज	थझ	थढ
द	दक	दख	दग	दज	दझ	दढ
ध	धक	धख	धग	धज	धझ	धढ
न	नक	नख	नग	नज	नझ	नढ
प	पक	पख	पग	पज	पझ	पढ

फ्	फ़क	फख	फग	फज	फड	फढ
ब्	ब़क	बख	बग	बज	बड	बढ
भ्	भ़क	भख	भग	भज	भड	भढ
म्	म़क	मख	मग	मज	मड	मढ
य्	य़क	यख	यग	यज	यड	यढ
र्	ऱक	रख	रग	रज	रड	रढ
ल्	ल़क	लख	लग	लज	लड	लढ
व्	व़क	वख	वग	वज	वड	वढ
श्	श़क	शख	शग	शज	शड	शढ
ष्	ष़क	षख	षग	षज	षड	षढ
स्	स़क	सख	सग	सज	सड	सढ
ह्	ह़क	हख	हग	हज	हड	हढ
क्	क़क	कख	कग	कज	कड	कढ
ख्	ख़क	खख	खग	खज	खड	खढ
ग्	ग़क	गख	गग	गज	गड	गढ
ज्	ज़क	जख	जग	जज	जड	जढ
फ़्	फ़क	फ़ख	फ़ग	फ़ज	फ़ड	फ़ढ
इ़्	इ़क	इ़ख	इ़ग	इ़ज	इ़ड	इ़ढ
ढ़्	ढ़क	ढ़ख	ढ़ग	ढ़ज	ढ़ड	ढ़ढ

### 6.3.3.2 CHCHC ( combination of three Consonants)

These are not as frequent as the CHC combinations. Only the major are listed below.

With a few exceptions these are mainly linear in nature. A majority of these are due to loan words which have entered Hindi. Combinations with rakar and rafar forms will not be considered. Similarly combinations with mātrās have not been taken into account.

क्द्र=इलेक्ट्रो	द्र्य=दारिद्र्य	थ्र्य=असामर्थ्य
क्ट्स=स्पैक्ट्स	द्व्य=द्व्यर्थक	र्द्ध=अर्द्ध
क्षण=तीक्ष्ण	न्त्स=स्ट्रडन्त्स	र्द्र=आर्द्रता
क्ष्म=भाग्यलक्ष्मी	न्ध=अनिन्ध	र्द्ध=अंतर्द्ध
क्ष्य=अंतःसाक्ष्य	न्द्र=उपेन्द्रनाथ	थ्र्य=अंतर्ध्यान
क्स्ड=फिक्स्ड	फ्ट्स=ड्राफ्ट्समैनौ	ध्व=उध्व
ग्भ=दिग्भ्रमित	र्क्त=पैमुक्तये	पर्य=कपर्यू
च्छव=उच्छवास	क्य=अतक्य	र्भ=निर्भ्रात
ज्ज्व=उज्ज्वल	र्क्र=अंतर्क्रिया	र्म्य=साधर्म्य
ण्ट्स=सर्वेण्ट्स	क्र्स=माक्र्स	लर्ड=फ्रीवर्ड
त्क्ष=तत्क्षण	घ्य=अघ्य	लर्स=चालर्स
त्प=उत्पत्ति	च्छ=मूर्च्छित	व्य=चतुर्व्यूह
त्व=तत्त्व	र्ज्ञ=अंतर्ज्ञान	श्र्य=पार्श्व
त्प्र=उत्प्रेरक	ज्य=वज्य	ष्य=ईर्ष्या
त्म्य=तादात्म्य	ट्र=पोर्ट्रेट	स्र्क=लुनाचास्की
त्र्य=वैचित्र्य	टर्स=आटर्स	स्ट=फस्ट
त्स्फ=उत्स्फूर्त	इर्स=वडर्सवर्थ	ल्व्य=केल्व्यूलेशन
त्स्य=मत्स्य	र्त=अमूर्त	ल्ट्र=पोल्ट्रीफार्म
दग्=उदग्ग्रीव	र्त्र=सारत्र	ल्ट्र=चिल्ट्रन
द्भ=उद्भ्रांतावस्था	त्र्स=भर्त्सना	लस्त=तोल्स्तोय

ष्क्र=निष्क्रिय

स्क्र=स्क्रिप्ट

स्थय=अस्वास्थ्य

ष्टय=विशिष्टय

स्क्व=मस्क्वा

स्प=स्प्रिंगदार

ष्ट्र=अंतरराष्ट्रीय

स्ट्र=आर्केस्ट्रा

स्प्ल=एस्प्लेनेड

ष्प्र=निष्प्रयोजन

स्त्र=अधोवस्त्र

#### 6.3.3.3.CHCHCHC ( Combination of four Consonants)

This is rare and is found only in the following which are English words transliterated into Devanagari.:

न्स्ट्रः कन्स्ट्रक्षन <sup>24</sup>
-------------------------------------

---

<sup>24</sup> Cited in Wikimapia and in <http://navbharattimes.indiatimes.com/new-tension-in-noida-extension/articleshow/15305178.cms>

### 6.3.4 The Collation Order of Hindi.

Collation is one of the most important features of a script grammar. It determines the order in which a given culture indexes its characters. This is best seen in a dictionary sort where for easy search words are sorted and arranged in a specific order. Within a given script, each allo-script may have a different sort-order. Thus in Hindi the conjunct glyph क्ष is sorted along with क, since the first letter of that conjunct is क and on a similar principle ज्ञ is sorted along with ज. The same is not the case with Marathi and Nepali which admit a different sort order.

Different scripts admit different sort orders and for all high-end NLP applications, sort is a crucial feature to ensure that the applications index data as per the cultural perception of that community. In quite a few States, sort order is clearly defined by the statutory bodies of that state and hence it is crucial that such sort order be ascertained and introduced in the script grammar.

In the case of Hindi the following is the traditional sort order as determined by the experts and used in dictionaries.

अ	आ	इ	ई	उ	ऊ	ऋ	ए	ऐ <sup>25</sup>	ऐ	ओ	औ
औ	क	ख	ग	घ	ङ	च	छ	ज	झ	ञ	ट
ठ	ड	ढ	ण	त	थ	द	ध	न	प	फ	ब
भ	म	य	र	ल	व	श	ष	स	ह		

The order as given below is pertinent to sorting by a computer program and is compliant with CLDR as laid down by Unicode and W3C.

ँ	ं	ः	अ	आ	इ	ई	उ	ऊ	ऋ	ए	ऐ
ऐ	ओ	औ	औ	क	ख	ग	घ	ङ	च	छ	ज
झ	ञ	ट	ठ	ड	ढ	ण	त	थ	द	ध	न
प	फ	ब	भ	म	य	र	ल	व	श	ष	स
ह	ा	ि	ी	ु	ू	ृ	े	ँ	ै	ो	ौ
ौ	्	ं									

---

<sup>25</sup> ऐ and औ are normally not listed in traditional dictionaries, but are included as digital dictionaries do include these characters.



In Tabular format:

ँ	ं	ः	अ	आ	इ	ई	उ	ऊ	ऋ	ए	ऐ
ऐ	ओ	औ	औ	क	ख	ग	घ	ङ	च	छ	ज
झ	ञ	ट	ठ	ड	ढ	ण	त	थ	द	ध	न
प	फ	ब	भ	म	य	र	ल	व	श	ष	स
ह	ा	ि	ी	ु	ू	ृ	े	ँ	ै	ो	ॉ
ौ	्	ं									

## 7. REFERENCES

1. <http://www.unicode.org>
2. ISCII'91
3. केंद्रीय हिंदी निदेशालय: देवनागरी लिपि तथा हिंदी वर्तनी का मानकीकरण नई दिल्ली. 2006

## 8. ANNEXURES

### Annexure 1: Names of experts who have contributed to the script grammar

➤ **Prepared By :**

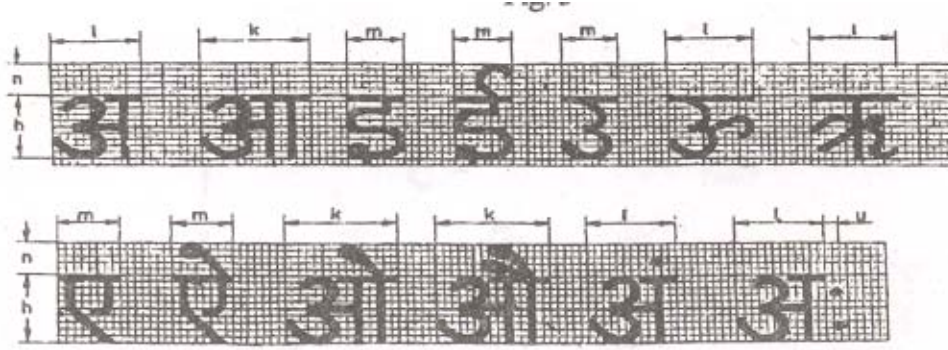
- ✓ Dr. Raimond Doctor, Consultant, C-DAC, GIST, Pune.
- ✓ Shapes in this document are based on CHD book.

➤ **Inputs By :**

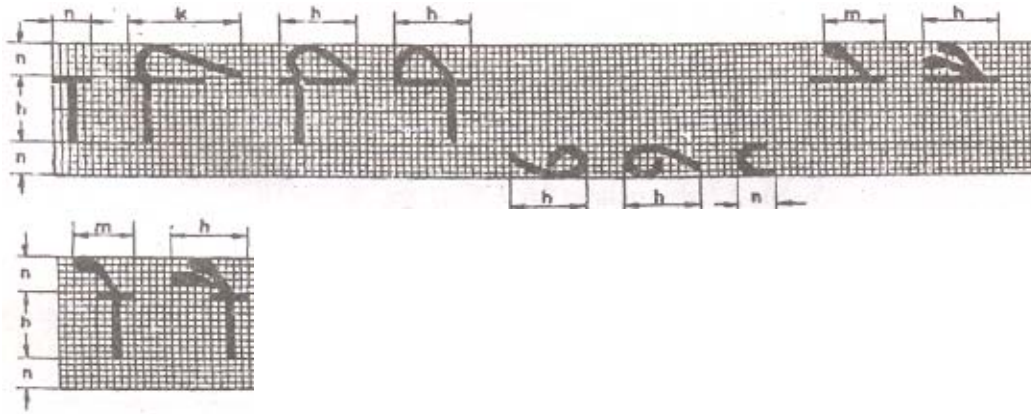
- ✓ Mr. Prabhakar Pandey, Hindi Officer, C-DAC, Pune.
- ✓ Mr. Tukaram Patil, Department of Hindi.
- ✓ Dr. Sudhir Mishra, Hindi Linguist, C-DAC, Pune.
- ✓ Hindi Language expert nominated by, Director, CIIL ( Dr. R. Sachdeva)

## Annexure 2: Shapes of Hindi characters as per Central Hindi Directorate

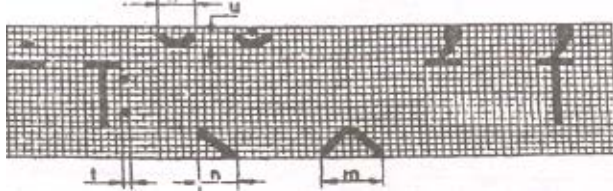
### Hindi Vowels



### Hindi Vowel Modifiers: Matras



### Hindi Diacritics and Rafar(Repha) Rakar



## Hindi Consonants

क ख ग घ ङ

ख

च छ ज झ ञ

ज

ट ठ ड ढ ण

ड ढ

त थ द ध न

प फ ब भ म

फ

य र ल व

व

श ष स ह

## Ligatural Representations

क्ष त्र ज्ञ श्र

### Annexure 3: Unicode Table of Devanāgarī <sup>26</sup>

---

<sup>26</sup> The Unicode chart provided is for version 5.1 since the Script Grammar was prepared at that time. No considerable change in the script grammar can be seen in the updated versions of Unicode, with the possible addition of the Rupee Sign U+02B9.