

**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.....	18
6.3.1. The Character Set of Hindi. ....	18
6.3.2. Consonant Mātrā Combinations. ....	25
6.3.3. The Ligature Set of Hindi. ....	39
6.3.4 The Collation Order of Hindi.....	49
7. REFERENCES .....	51
8. ANNEXURES .....	52
Annexure 1: Names of experts who have contributed to the script grammar.....	52
Annexure 2: Shapes of Hindi characters as per Central Hindi Directorate .....	54
Annexure 3: Unicode Table of Devanāgarī .....	56

## 0. INTRODUCTION

The term **script grammar** refers to the behaviour 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 history of the script and language (wherever available)
- the syllabic structure of the writing system of the language
- the rule ordering of the characters within the syllable (akshar)
- description of the syllabic clusters
- 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 tries 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 Hindi although both languages share the same script. Thus in Hindi क्ष, ञ 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 ञ 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.

- Most importantly it 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, 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.

**Allo-Script:** 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 ङ [U+095C] ढ [U+095D] are not used in Konkani. These sub-sets of scripts based on a single “matricial” script are termed as allo-scripts.

**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 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 Indic 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, the one or more dead

---


<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

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 breath is at least partly obstructed.

**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 // in Devanāgarī is an example of a displaced catenator.

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

**Display rendition:** 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, Hindi and Marathi. It is treated as different from the र̣ (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.

**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 breath passing through the nose.  
Example *m n* in English.

**Nasalized vowel:** A vowel pronounced with the breath 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 placed a nasal 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 कर्क , र्+घ्+य घ्य

**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 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 dental 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

**Vowel sign:** A graphic character associated with a letter, to indicate 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 allo-scripts 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 क्त 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 syllable (akshar) structure described above. It can admit up to 3 consonant clusters.**

Alphasyllabaries are determined by the notion of the syllable or 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 syllable (akshar), first defined in the ISCII document (1991), identifies the following character 'sub-sets' for the purposes of identifying the syllable (akshar). In what follows the syllable analysis will be restricted to Hindi.

### (C) Consonants

क	क़	ख	ख़	ग
ग़	घ	ङ	च	छ
ज	ज़	झ	ञ	ट
ठ	ड	ड़	ढ	ढ़
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	ओ	औ	ँ	ऑ
---	---	---	---	---	---	---	---	---	---	---	---	---

### (M) Mātrās or Vowel Modifiers

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

### (D) Diacritics

◌ं: Anuswar	Anuswara, a nasal, 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 above the other.

ऽ: Avagraha <sup>5</sup>	For extra length with long vowels as seen in the Sanskrit text /उपदेशेऽजनुनासिक इत् /
--------------------------	---

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

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

Each of these sub-types has its restrictions in terms of what can precede or follow it, within a syllable (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 syllable (akshar).

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

### 6.2.1.Syllable (akshar) Types

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

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

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

1. The Vowel syllable (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 क, ख, ग, ज, ङ

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

2. The Consonant syllable (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:/, क्री /qi:/

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

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

2.5. A consonant cluster i.e. a dead or 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 /tkã/, त्कः /tkaH/ त्कुं /tkũ/, त्दु /tdu/.

The above permutations and combinations result in 7 major syllable (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, four consonants can come together: कार्त्स्न्य /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	CHCHCHCM

A total number of 16 theoretical syllables is therefore possible. It will be seen that the written syllable (akshar) is not very different in structure from the phonetic syllable and that the movement from the written to the spoken levels is made feasible by application of certain rules.

<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



Since the formal structure script grammar of the syllable (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.

## 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 - aspirated
Pharyngal	क <sup>11</sup>	ख	ग <sup>9</sup>			
Velar	क	ख	ग	घ	ङ	
Palatal	च	छ	ज	झ	ञ	ज़
Retroflex	ट	ठ	ड	ढ	ण	

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

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

Flaps			ड़	ढ़		
Dental	त	थ	द	ध	न	
Labio-Dental						फ़
B-labial	प	फ	ब	भ	म	

#### Other consonants

य	र	ऱ	ल	व
श	ष	स	ह	

**Note:** Ligatures क्षत्रज्ञश्र 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 -/+aspirated
Pharyngeal	क़ <sup>13</sup>	ख़	ग़			
Velar	क	ख	ग	घ	ङ	
Palatal	च	छ	ज	झ	ञ	ज्ञ
Retroflex	ट	ठ	ड	ढ	ण	
Flaps			ड़	ढ़		
Dental	त	थ	द	ध	न	
Labio-Dental						फ़
B-labial	प	फ	ब	भ	म	

#### Other consonants

य	र	ऱ	ल	व
श	ष	स	ह	

<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> क़ग़ are Not recognized by the Central Hindi Directorate cf. p.2 of their document



### 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
ँ	DEVANAGARI LETTER CHANDRA E
ॉ	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 ओ औ ँ ऑ

### 6.3.1.3. The Mātrā Set

The Mātrā (Vowel Modifier Set) 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 (There can not be any matra on this)	क् + ए = के
8. Devanāgarī sign AI	ै	ऐ (There can not be any matra on this)	क् + ऐ = कै
9..Devanāgarī sign O	ो	ओ	क् + ओ = को
10. Devanāgarī sign AU	ौ	औ	क् + औ = कौ
11 Devanāgarī sign Candra E	ँ	ँ	क् + ँ = कँ
12 Devanāgarī sign Candra O	ॉ	ॉ	क् + ऌ = कॉ

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

ा ि ी ु ू ृ े ै ो ौ ँ ऌ

#### 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 left of character	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: कर्त्त e.g. top rafars will be formed in case of following words.

धर्म, चर्खा
-------------

#### RAKARS

1. Bottom rakar र्र र्र
2. Side rakar र्र र्र

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

Bottom rakar	र्रम, राष्ट्र
Side rakar	र्रत, र्रक, र्रसाद, र्रस्रोत

#### 6.3.1.6. Diacritics

These are as under in the case of Hindi:

ं : - Anuswar रंग

ँ : -Candrabindu/Anunasika र्रँट

् : - Halant र्रंवत्

ः - Visarga दुःख

s : - For extra length with long vowels e.g. / उपदेशेऽ जनुनासिक इत्/ <sup>14</sup>

### 6.3.1.7. Numerals

The International number set (Latino-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 language.

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 Deergha 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.

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 ॥

<sup>14</sup> Avagraha is rarely used in Hindi and is used for representing text in Sanskrit here the character is needed..



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 क्र ख ग ज़ फ़ ड़ ढ़

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.

e.g. Although the combination of ड़ +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: ड़ा

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	ङा	चा	छा	जा	झा	जा
ि	कि	खि	गि	घि	ङि	चि	छि	जि	झि	ञि
ी	की	खी	गी	घी	ङी	ची	छी	जी	झी	जी
ु	कु	खु	गु	घु	ङु	चु	छु	जु	झु	जु
ू	कू	खू	गू	घू	ङू	चू	छू	जू	झू	जू
ृ	कृ	खृ	गृ	घृ	ङृ	चृ	छृ	जृ	झृ	जृ
े	के	खे	गे	घे	ङे	चे	छे	जे	झे	जे
ै	कै	खै	गै	घै	ङै	चै	छै	जै	झै	जै
ो	को	खो	गो	घो	ङो	चो	छो	जो	झो	जो
ौ	कौ	खौ	गौ	घौ	ङौ	चौ	छौ	जौ	झौ	जौ
ँ	कँ	खँ	गँ	घँ	ङँ	चँ	छँ	जँ	झँ	जँ
ॉ	कॉ	खाँ	गाँ	घाँ	ङाँ	चाँ	छाँ	जाँ	झाँ	जाँ

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	णा	ता	था	दा	धा	ना
ि	टि	ठि	डि	ढि	णि	ति	थि	दि	धि	नि
ी	टी	ठी	डी	ढी	णी	ती	थी	दी	धी	नी
ु	टु	ठु	डु	ढु	णु	तु	थु	दु	धु	नु
ू	टू	ठू	डू	ढू	णू	तू	थू	दू	धू	नू
ृ	टृ	ठृ	डृ	ढृ	णृ	तृ	थृ	दृ	धृ	नृ
े	टे	ठे	डे	ढे	णे	ते	थे	दे	धे	ने
ै	टै	ठै	डै	ढै	णै	तै	थै	दै	धै	नै
ो	टो	ठो	डो	ढो	णो	तो	थो	दो	धो	नो
ौ	टौ	ठौ	डौ	ढौ	णौ	तौ	थौ	दौ	धौ	नौ
ँ	टँ	ठँ	डँ	ढँ	णँ	तँ	थँ	दँ	धँ	नँ
ॉ	टाँ	ठाँ	डाँ	ढाँ	णाँ	ताँ	थाँ	दाँ	धाँ	नाँ

### 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
ँ	पँ	फँ	बँ	भँ	मँ	यँ	रँ	लँ	वँ	शँ	षँ	सँ	हँ
ै	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
ॉ	पाँ	फाँ	बाँ	भाँ	माँ	याँ	राँ	लाँ	वाँ	शाँ	षाँ	साँ	हाँ

#### Consonant and Mātrā combinations Set 4

This set is in continuation of set 3 which shows consonant and Matra combinations for Nukta Consonants.<sup>15</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>16</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
ौ	कौ	खौ	गौ	जौ	फौ	डौ	ढौ
ँ	कँ	खँ	गँ	जँ	फँ	डँ	ढँ
ॉ	कॉ	खॉ	गॉ	जॉ	फॉ	डॉ	ढॉ

<sup>15</sup> क,ग are not recognized as part of the Consonant set by the Central Hindi Directorate

<sup>16</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	ञिं
ीं	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	जूं	झृं	जूं
ें	कें	खें	गें	घें	ङें	चें	छें	जें	झें	जें
ैं	कैं	खैं	गैं	घैं	ङैं	चैं	छैं	जैं	झैं	जैं
ों	कों	खों	गों	घों	ङों	चों	छों	जों	झों	जों
ौं	कौं	खौं	गौं	घौं	ङौं	चौं	छौं	जौं	झौं	जौं
ँ	कँ	खँ	गँ	घँ	ङँ	चँ	छँ	जँ	झँ	जँ
ाँ	काँ	खाँ	गाँ	घाँ	ङाँ	चाँ	छाँ	जाँ	झाँ	जाँ

## 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
ूं	टूं	ठूं	डूं	ढूं	णूं	तूं	थूं	दूं	धूं	नूं
ृं	ट्रं	ठ्रं	ड्रं	ढ्रं	ण्रं	त्रं	थ्रं	द्रं	ध्रं	न्रं
ें	टेँ	ठेँ	डेँ	ढेँ	णेँ	तेँ	थेँ	देँ	धेँ	नेँ
ैं	टैँ	ठैँ	डैँ	ढैँ	णैँ	तैँ	थैँ	दैँ	धैँ	नैँ
ों	टों	ठों	डों	ढों	णों	तों	थों	दों	धों	नों
ौं	टौं	ठौं	डौं	ढौं	णौं	तौं	थौं	दौं	धौं	नौं
ँ <sup>17</sup>	टँ	ठँ	डँ	ढँ	णँ	तँ	थँ	दँ	धँ	नँ
ॉं	टाँ	ठाँ	डाँ	ढाँ	णाँ	ताँ	थाँ	दाँ	धाँ	नाँ

<sup>17</sup> Since ँ ,ँॉ are used only for the purpose of writing loan words mainly from English and since English does not have nasals, 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
ूं	पूं	फूं	बूं	भूं	मूं	यूं	रूं	लूं	वूं	शूं	षूं	सूं	हूं
ृं	पृं	फृं	बृं	भृं	मृं	यृं	X	लृं	वृं	शृं	षृं	सृं	हृं
ें	पें	फें	बें	भें	में	यें	रें	लें	वें	शें	षें	सें	हें
ैं	पैं	फैं	बैं	भैं	मैं	यैं	रैं	लैं	वैं	शैं	षैं	सैं	हैं
ों	पों	फों	बों	भों	मों	यों	रों	लों	वों	शों	षों	सों	हों
ौं	पौं	फौं	बौं	भौं	मौं	यौं	रौं	लौं	वौं	शौं	षौं	सौं	हौं
ँ <sup>18</sup>	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

<sup>18</sup> Since ँ ,ॉ are used only for the purpose of writing loan words mainly from English and since English does not have nasals, 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
ूं	कू	खू	गू	जू	फू	डू	ढू
ृं	कृ	खृ	गृ	जू	फृ	डृ	ढृ
ें	के	खे	गे	जे	फे	डे	ढे
ैं	कै	खै	गै	जै	फै	डै	ढै
ो	को	खो	गो	जो	फो	डो	ढो
ौ	कौ	खौ	गौ	जौ	फौ	डौ	ढौ
िं	किं	खिं	गिं	जिं	फिं	डिं	ढिं
ँ <sup>19</sup>	कँ	खँ	गँ	जँ	फँ	डँ	ढँ
ॉ	काँ	खाँ	गाँ	जाँ	फाँ	डाँ	ढाँ

<sup>19</sup> Since ँ ,ँ are used only for the purpose of writing loan words mainly from English and since English does not have nasals, the nasal shapes of ँ ,ँ are rarely encountered in Hindi

**Consonant and Mātrā + Nasal combinations: With Chandrabindu<sup>20</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	जूँ	झृँ	जूँ
ेँ	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
ौँ	कोँ	खौँ	गौँ	घौँ	ङौँ	चौँ	छौँ	जौँ	झौँ	जौँ

<sup>20</sup> As per rule 2.6.2.3. of the Central Hindi Directorate, Chandrabindu cannot be placed over matras placed above the Shirrekha 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
ूँ	टूँ	ठूँ	डूँ	ढूँ	णूँ	तूँ	थूँ	दूँ	धूँ	नूँ
्रूँ	ट्रूँ	ठ्रूँ	ड्रूँ	ढ्रूँ	ण्रूँ	त्रूँ	थ्रूँ	द्रूँ	ध्रूँ	न्रूँ
ृँ	टृँ	ठृँ	डृँ	ढृँ	णृँ	तृँ	थृँ	दृँ	धृँ	नृँ
ेँ	टेँ	ठेँ	डेँ	ढेँ	णेँ	तेँ	थेँ	देँ	धेँ	नेँ
ैँ	टैँ	ठैँ	डैँ	ढैँ	णैँ	तैँ	थैँ	दैँ	धैँ	नैँ
ोँ <sup>21</sup>	टोँ	ठोँ	डोँ	ढोँ	णोँ	तोँ	थोँ	दोँ	धोँ	नोँ
ौँ	टौँ	ठौँ	डौँ	ढौँ	णौँ	तौँ	थौँ	दौँ	धौँ	नौँ

<sup>21</sup>Since ँ, ्रूँ are used only for the purpose of writing loan words mainly from English and since English does not have nasals, the nasal shapes of ँ, ्रूँ are rarely encountered in Hindi and are therefore not included.

### 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
ूँ	पूँ	फूँ	बूँ	भूँ	मूँ	यूँ	रूँ	लूँ	वूँ	शूँ	षूँ	सूँ	हूँ
्रूँ	प्रूँ	फ्रूँ	ब्रूँ	भ्रूँ	म्रूँ	य्रूँ	र्रूँ	ल्रूँ	व्रूँ	श्रूँ	ष्रूँ	स्रूँ	ह्रूँ
ृँ	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
ैँ	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
ौँ	पौँ	फौँ	बौँ	भौँ	मौँ	यौँ	रौँ	लौँ	वौँ	शौँ	षौँ	सौँ	हौँ

### 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	ब	व
ँ	कँ	खँ	गँ	जँ	फँ	बँ	वँ
ाँ	काँ	खाँ	गाँ	जाँ	फाँ	बाँ	वाँ
िँ	किँ	खिँ	गिँ	जिँ	फिँ	बिँ	विँ
ीँ	कीँ	खीँ	गीँ	जीँ	फीँ	बीँ	वीँ
ुँ	कुँ	खुँ	गुँ	जुँ	फुँ	बुँ	वुँ
ूँ	कूँ	खूँ	गूँ	जूँ	फूँ	बूँ	वूँ
ृँ	कृँ	खृँ	गृँ	जृँ	फृँ	बृँ	वृँ
ेँ	केँ	खेँ	गेँ	जेँ	फेँ	बेँ	वेँ
ैँ	कैँ	खैँ	गैँ	जैँ	फैँ	बैँ	वैँ
ोँ	कोँ	खोँ	गोँ	जोँ	फोँ	बोँ	वोँ
ौँ	कौँ	खौँ	गौँ	जौँ	फौँ	बौँ	वौँ
ँ	कँ	खँ	गँ	जँ	फँ	बँ	वँ
ाँ	काँ	खाँ	गाँ	जाँ	फाँ	बाँ	वाँ

### 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 eg. 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	क्क	क्ख	क्ग	क्घ	क्ङ	क्च	क्छ	क्ज	क्झ	क्ञ
ग्	क्क	क्ख	क्ग	क्घ	क्ङ	क्च	क्छ	क्ज	क्झ	क्ञ
घ्	क्क	क्ख	क्ग	क्घ	क्ङ	क्च	क्छ	क्ज	क्झ	क्ञ
ङ्	क्क	क्ख	क्ग	क्घ	क्ङ	क्च	क्छ	क्ज	क्झ	क्ञ
च्	क्क	क्ख	क्ग	क्घ	क्ङ	क्च	क्छ	क्ज	क्झ	क्ञ
छ्	क्क	क्ख	क्ग	क्घ	क्ङ	क्च	क्छ	क्ज	क्झ	क्ञ
ज्	क्क	क्ख	क्ग	क्घ	क्ङ	क्च	क्छ	क्ज	क्झ	क्ञ
झ्	क्क	क्ख	क्ग	क्घ	क्ङ	क्च	क्छ	क्ज	क्झ	क्ञ
ञ्	क्क	क्ख	क्ग	क्घ	क्ङ	क्च	क्छ	क्ज	क्झ	क्ञ
ट्	क्क	क्ख	क्ग	क्घ	क्ङ	क्च	क्छ	क्ज	क्झ	क्ञ
ठ्	क्क	क्ख	क्ग	क्घ	क्ङ	क्च	क्छ	क्ज	क्झ	क्ञ
ड्	क्क	क्ख	क्ग	क्घ	क्ङ	क्च	क्छ	क्ज	क्झ	क्ञ
ढ्	क्क	क्ख	क्ग	क्घ	क्ङ	क्च	क्छ	क्ज	क्झ	क्ञ

ण्	णक	णख	णग	णघ	णङ	णच	णछ	णज	णझ	णञ
त्	त्क	त्ख	त्ग	त्घ	त्ङ	त्व	त्छ	त्ज	त्झ	त्ञ
थ्	थक	थख	था	थघ	थङ	थच	थछ	थज	थझ	थञ
द्	द्क	द्ख	द्ग	द्घ	द्ङ	द्च	द्छ	द्ज	द्झ	द्ञ
ध्	धक	धख	धा	धघ	धङ	धच	धछ	धज	धझ	धञ
न्	न्क	न्ख	ना	न्घ	न्ङ	न्च	न्छ	न्ज	न्झ	न्ञ
प्	पक	पख	पा	पघ	पङ	पच	पछ	पज	पझ	पञ
फ्	फक	फख	फग	फघ	फङ	फच	फछ	फज	फझ	फञ
ब्	बक	बख	बा	बघ	बङ	बच	बछ	बज	बझ	बञ
भ्	भक	भख	भा	भघ	भङ	भच	भछ	भज	भझ	भञ
म्	मक	मख	मा	मघ	मङ	मच	मछ	मज	मझ	मञ
य्	यक	यख	या	यघ	यङ	यच	यछ	यज	यझ	यञ
र्	र्क	र्ख	र्ग	र्घ	र्ङ	र्च	र्छ	र्ज	र्झ	र्ञ
ल्	लक	लख	ला	लघ	लङ	लच	लछ	लज	लझ	लञ
व्	वक	वख	वा	वघ	वङ	वच	वछ	वज	वझ	वञ
श्	शक	शख	शग	शघ	शङ	शच	शछ	शज	शझ	शञ
ष्	षक	षख	षा	षघ	षङ	षच	षछ	षज	षझ	षञ
स्	सक	सख	सा	सघ	सङ	सच	सछ	सज	सझ	सञ
ह्	हक	हख	हा	हघ	हङ	हच	हछ	हज	हझ	हञ
क्	कक	कख	का	कघ	कङ	कच	कछ	कज	कझ	कञ
ख्	खक	खख	खा	खघ	खङ	खच	खछ	खज	खझ	खञ
ग्	गक	गख	गा	गघ	गङ	गच	गछ	गज	गझ	गञ
ज्	जक	जख	जा	जघ	जङ	जच	जछ	जज	जझ	जञ
फ्	फक	फख	फा	फघ	फङ	फच	फछ	फज	फझ	फञ
ड्	डक	डख	डा	डघ	डङ	डच	डछ	डज	डझ	डञ
ढ्	ढक	ढख	ढा	ढघ	ढङ	ढच	ढछ	ढज	ढझ	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 eg. 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	कन
ख्	खट	खठ	खड	खढ	खण	ख्त	खथ	खद	खध	खन
ग्	गट	गठ	गड	गढ	गण	गत	गथ	गद	गध	गन
घ्	घट	घठ	घड	घढ	घण	घ्त	घथ	घद	घध	घन
ङ्	ङट	ङठ	ङड	ङढ	ङण	ङ्त	ङथ	ङद	ङध	ङन
च्	चट	चठ	चड	चढ	चण	च्त	चथ	चद	चध	चन
छ्	छट	छठ	छड	छढ	छण	छ्त	छथ	छद	छध	छन
ज्	जट	जठ	जड	जढ	जण	ज्त	जथ	जद	जध	जन
झ्	झट	झठ	झड	झढ	झण	झ्त	झथ	झद	झध	झन
ञ्	ञट	ञठ	ञड	ञढ	ञण	ञ्त	ञथ	ञद	ञध	ञन
ट्	टट	टठ	टड	टढ	टण	ट्त	टथ	टद	टध	टन
ठ्	ठट	ठठ	ठड	ठढ	ठण	ठ्त	ठथ	ठद	ठध	ठन
ड्	डट	डठ	डड	डढ	डण	ड्त	डथ	डद	डध	डन
ढ्	ढट	ढठ	ढड	ढढ	ढण	ढ्त	ढथ	ढद	ढध	ढन
ण्	णट	णठ	णड	णढ	णण	ण्त	णथ	णद	णध	णन
त्	त्ट	त्ठ	त्ड	त्ढ	त्ण	त्त	त्थ	त्द	त्ध	त्न
थ्	थट	थठ	थड	थढ	थण	थ्त	थथ	थद	थध	थन
द्	द्दट	द्दठ	द्दड	द्दढ	द्दण	द्दत्त	द्दथ	द्दद	द्दध	द्दन
ध्	धट	धठ	धड	धढ	धण	ध्त	धथ	धद	धध	धन
न्	न्ट	न्ठ	न्ड	न्ढ	न्ण	न्त	न्थ	न्द	न्ध	न्न
प्	पट	पठ	पड	पढ	पण	प्त	पथ	पद	पध	पन
फ्	फट	फठ	फड	फढ	फण	फ्त	फथ	फद	फध	फन
ब्	बट	बठ	बड	बढ	बण	ब्त	बथ	बद	बध	बन
भ्	भट	भठ	भड	भढ	भण	भ्त	भथ	भद	भध	भन
म्	मट	मठ	मड	मढ	मण	म्त	मथ	मद	मध	मन
य्	यट	यठ	यड	यढ	यण	य्त	यथ	यद	यध	यन

र	र	र	र	र	र	र	र	र	र	र
ल	ल	ल	ल	ल	ल	ल	ल	ल	ल	ल
व	व	व	व	व	व	व	व	व	व	व
श	श	श	श	श	श	श	श	श	श	श
ष	ष	ष	ष	ष	ष	ष	ष	ष	ष	ष
स	स	स	स	स	स	स	स	स	स	स
ह	ह	ह	ह	ह	ह	ह	ह	ह	ह	ह
क	क	क	क	क	क	क	क	क	क	क
ख	ख	ख	ख	ख	ख	ख	ख	ख	ख	ख
ग	ग	ग	ग	ग	ग	ग	ग	ग	ग	ग
ज	ज	ज	ज	ज	ज	ज	ज	ज	ज	ज
फ	फ	फ	फ	फ	फ	फ	फ	फ	फ	फ
ड	ड	ड	ड	ड	ड	ड	ड	ड	ड	ड
ढ	ढ	ढ	ढ	ढ	ढ	ढ	ढ	ढ	ढ	ढ

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 eg. Combination of consonant “क” and “प” is the ligature “कप”.

CHC( combination of two consonants) - Set 3													
	प	फ	ब	भ	म	य	र	ल	व	श	ष	स	ह
क्	कप	कफ	कब	कभ	कम	क्य	क्र	क्ल	क्व	कश	कष	कस	कह
ख्	खप	खफ	खब	खभ	खम	ख्य	ख्र	ख्ल	ख्व	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	घ्ल	घ्व	घश	घष	घस	घह
ङ्	ङप	ङफ	ङब	ङभ	ङम	ङ्य	ङ्र	ङ्ल	ङ्व	ङश	ङष	ङस	ङह
च्	चप	चफ	चब	चभ	चम	च्य	च्र	च्ल	च्व	चश	चष	चस	चह
छ्	छप	छफ	छब	छभ	छम	छ्य	छ्र	छ्ल	छ्व	छश	छष	छस	छह
ज्	जप	जफ	जब	जभ	जम	ज्य	ज्र	ज्ल	ज्व	जश	जष	जस	जह
झ्	झप	झफ	झब	झभ	झम	झ्य	झ्र	झ्ल	झ्व	झश	झष	झस	झह
ञ्	ञप	ञफ	ञब	ञभ	ञम	ञ्य	ञ्र	ञ्ल	ञ्व	ञश	ञष	ञस	ञह
ट्	टप	टफ	टब	टभ	टम	ट्य	ट्र	ट्ल	ट्व	टश	टष	टस	टह
ठ्	ठप	ठफ	ठब	ठभ	ठम	ठ्य	ठ्र	ठ्ल	ठ्व	ठश	ठष	ठस	ठह
ड्	डप	डफ	डब	डभ	डम	ड्य	ड्र	ड्ल	ड्व	डश	डष	डस	डह
ढ्	ढप	ढफ	ढब	ढभ	ढम	ढ्य	ढ्र	ढ्ल	ढ्व	ढश	ढष	ढस	ढह
ण्	णप	णफ	णब	णभ	णम	ण्य	ण्र	ण्ल	ण्व	णश	णष	णस	णह
त्	तप	तफ	तब	तभ	तम	त्य	त्र	त्ल	त्व	तश	तष	तस	तह
थ्	थप	थफ	थब	थभ	थम	थ्य	थ्र	थ्ल	थ्व	थश	थष	थस	थह
द्	दप	दफ	दब	दभ	दम	द्व	द्र	द्ल	द्व	दश	दष	दस	दह
ध्	धप	धफ	धब	धभ	धम	ध्व	ध्र	ध्ल	ध्व	धश	धष	धस	धह
न्	नप	नफ	नब	नभ	नम	न्य	न्र	न्ल	न्व	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	फ्ल	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	भ्व	भश	भष	भस	भह
म्	मप	मफ	मब	मभ	मम	म्य	म्र	म्ल	म्व	मश	मष	मस	मह
य्	यप	यफ	यब	यभ	यम	य्य	य्र	य्ल	य्व	यश	यष	यस	यह
र्	रप	रफ	रब	रभ	रम	र्य	र्र	र्ल	र्व	रश	रष	रस	रह

ल्	ल्प	ल्फ	ल्ब	ल्भ	ल्म	ल्य	ल्र	ल्ल	ल्व	ल्श	ल्ष	ल्स	ल्ह
व्	वप	वफ	वब	वभ	वम	व्य	व्र	व्ल	व्व	व्श	व्ष	व्स	व्ह
श्	श्प	श्फ	श्ब	श्भ	श्म	श्य	श्च <sup>22</sup>	श्ल	श्व	श्श	श्ष	श्स	श्ह
ष्	ष्प	ष्फ	ष्व	ष्भ	ष्म	ष्य	ष्	ष्ल	ष्व	ष्श	ष्ष	ष्स	ष्ह
स्	स्प	स्फ	स्ब	स्भ	स्म	स्य	स्त्र	स्ल	स्व	स्श	स्ष	स्स	स्ह
ह्	हप	हफ	हब	हभ	हम	ह्य	ह्र	ह्ल	ह्व	ह्श	ह्ष	ह्स	ह्ह
क्	क्प	क्फ	क्ब	क्भ	क्म	क्य	क्	क्ल	क्व	क्श	क्ष	क्स	क्ह
ख्	खप	खफ	ख्व	खभ	खम	ख्य	ख्र	ख्ल	ख्व	ख्श	ख्ष	ख्स	ख्ह
ग्	गप	गफ	ग्व	गभ	गम	ग्य	ग्र	ग्ल	ग्व	ग्श	ग्ष	ग्स	ग्ह
ज्	जप	जफ	ज्व	जभ	जम	ज्य	ज्र	ज्ल	ज्व	ज्श	ज्ष	ज्स	ज्ह
फ्	फप	फफ	फ्व	फभ	फम	फ्य	फ्र	फल	फ्व	फ्श	फ्ष	फ्स	फ्ह
ड्	डप	डफ	ड्व	डभ	डम	ड्य	ड्र	ड्ल	ड्व	ड्श	ड्ष	ड्स	ड्ह
ढ्	ढप	ढफ	ढ्व	ढभ	ढम	ढ्य	ढ्र	ढ्ल	ढ्व	ढ्श	ढ्ष	ढ्स	ढ्ह

<sup>22</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 forms, select one consonant from the first column and second from first row. For eg. Combination of consonant “क” and “ऋ” is ligature “क्क़”.

CHC Set 4						
	क	ख	ग	ज	ड़	ढ़
क्	क्क़	क्ख़	क्ग़	क्ज़	क्ड़	क्ढ़
ख्	क्क़	क्ख़	क्ग़	क्ज़	क्ड़	क्ढ़
ग्	क्क़	क्ख़	क्ग़	क्ज़	क्ड़	क्ढ़
घ्	क्क़	क्ख़	क्ग़	क्ज़	क्ड़	क्ढ़
ङ्	ङ्क़	ङ्ख़	ङ्ग़	ङ्ज़	ङ्ड़	ङ्ढ़
च्	क्क़	क्ख़	क्ग़	क्ज़	क्ड़	क्ढ़
छ्	क्क़	क्ख़	क्ग़	क्ज़	क्ड़	क्ढ़
ज्	क्क़	क्ख़	क्ग़	क्ज़	क्ड़	क्ढ़
झ्	क्क़	क्ख़	क्ग़	क्ज़	क्ड़	क्ढ़
ञ्	क्क़	क्ख़	क्ग़	क्ज़	क्ड़	क्ढ़
ट्	क्क़	क्ख़	क्ग़	क्ज़	क्ड़	क्ढ़
ठ्	क्क़	क्ख़	क्ग़	क्ज़	क्ड़	क्ढ़
ड्	क्क़	क्ख़	क्ग़	क्ज़	क्ड़	क्ढ़
ढ्	क्क़	क्ख़	क्ग़	क्ज़	क्ड़	क्ढ़
ण्	क्क़	क्ख़	क्ग़	क्ज़	क्ड़	क्ढ़
त्	क्क़	क्ख़	क्ग़	क्ज़	क्ड़	क्ढ़
थ्	क्क़	क्ख़	क्ग़	क्ज़	क्ड़	क्ढ़
द्	क्क़	क्ख़	क्ग़	क्ज़	क्ड़	क्ढ़
ध्	क्क़	क्ख़	क्ग़	क्ज़	क्ड़	क्ढ़
न्	क्क़	क्ख़	क्ग़	क्ज़	क्ड़	क्ढ़
प्	क्क़	क्ख़	क्ग़	क्ज़	क्ड़	क्ढ़
फ्	क्क़	क्ख़	क्ग़	क्ज़	क्ड़	क्ढ़

ब्	ब्र	ब्र	बा	ब्र	ब्र	ब्र
भ्	भ्र	भ्र	भा	भ्र	भ्र	भ्र
म्	म्र	म्र	मा	म्र	म्र	म्र
य्	य्र	य्र	या	य्र	य्र	य्र
र्	र्र	र्र	रा	र्र	र्र	र्र
ल्	ल्र	ल्र	ला	ल्र	ल्र	ल्र
व्	व्र	व्र	वा	व्र	व्र	व्र
श्	श्र	श्र	शा	श्र	श्र	श्र
ष्	श्र	श्र	शा	श्र	श्र	श्र
स्	स्र	स्र	सा	स्र	स्र	स्र
ह्	ह्र	ह्र	हा	ह्र	ह्र	ह्र
क्	क्र	क्र	का	क्र	क्र	क्र
ख्	ख्र	ख्र	खा	ख्र	ख्र	ख्र
ग्	ग्र	ग्र	गा	ग्र	ग्र	ग्र
ज्	ज्र	ज्र	जा	ज्र	ज्र	ज्र
फ्	फ्र	फ्र	फा	फ्र	फ्र	फ्र
ड्	ड्र	ड्र	डा	ड्र	ड्र	ड्र
ढ्	ढ्र	ढ्र	ढा	ढ्र	ढ्र	ढ्र

### 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.

क्त्र=इलेक्ट्रोन	द्व्य=द्व्यर्थक	र्द्र=आर्द्रता
क्त्स=स्पैक्त्स	न्त्स=स्टूडन्त्स	द्वर्=अंतद्वर्द्व
क्ष्ण=तीक्ष्ण	न्द्य=अनिन्द्य	ध्व=अंतध्वान
क्ष्म=भाग्यलक्ष्मी	न्द्र=उपेन्द्रनाथ	ध्व=उध्व
क्ष्य=अंतःसाक्ष्य	फ्ट्स=झाफ्ट्समैनों	फ्र्य=कफ्र्यू
क्स्ड=फिक्स्ड	कर्त्त=पैमुकर्त्तये	भ्र्=निभ्र्नात
ग्भ्र=दिग्भ्रमित	कर्व्य=अतकर्व्य	म्र्य=साधम्र्य
च्छ्व=उच्छ्वास	र्क्र=अंतर्क्रिया	लर्ड=फ्रीवर्ल्ड
ज्ज्व=उज्ज्वल	कर्स=माकर्स	लर्स=चार्लर्स
ण्ट्स=सर्वेण्ट्स	घ्व=अघ्व	व्य=चतुर्व्यूह
त्क्ष=तत्क्षण	चर्छ=मूर्च्छित	श्व=पार्श्व
त्प=उत्पत्ति	र्ज्ञ=अंतर्ज्ञान	ष्व=ईर्ष्या
त्त्व=तत्त्व	जर्व्य=वर्ज्य	स्र्क=लुनाचास्की
त्प्र=उत्प्रेरक	र्द्र=पोर्द्रेट	स्र्ट=फर्स्ट
त्त्य=तादात्म्य	टर्स=आटर्स	ल्व्य=केल्व्यूलेशन
त्र्य=वैचित्र्य	डर्स=वडर्सवर्थ	ल्ट्र=पोल्ट्रीफार्म
त्स्फ=उत्स्फूर्त	र्त्त=अमूर्त्त	ल्ट्र=चिल्ट्रिन
त्स्य=मत्स्य	र्त्र=सार्त्र	लस्त=तोल्स्तोय
द्व्ग=उद्व्गीव	त्स=भर्त्सना	ष्क्र=निष्क्रिय
द्व्भ्र=उद्व्भ्रंतावस्था	थ्व्य=असामथ्व्य	ष्ट्य=विशिष्ट्य
द्व्य=दारिद्र्य	द्वर्ध=अद्वर्ध	ष्ट्र=अंतरराष्ट्रीय

ष्र=निष्रयोजन

स्क्र=स्क्रिप्ट

स्क्व=मस्क्वा

स्ट्र=आर्केस्ट्रा

स्त्र=अधोवस्त्र

स्थ्य=अस्वास्थ्य

स्प्र=सिंगदार

स्प्ल=एस्प्लेनेड

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

This is rare and is found only in the following:

न्स्ट्र: कन्स्ट्रक्षन, कन्स्ट्रक्ट



### 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>23</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	ा	ि	ी	ु	ू	ृ	े	ँ	ै	ो	ॉ
ौ	्	्र									

<sup>23</sup> ऐ and औ are normally not listed in traditional dictionaries, but are included as some modern 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

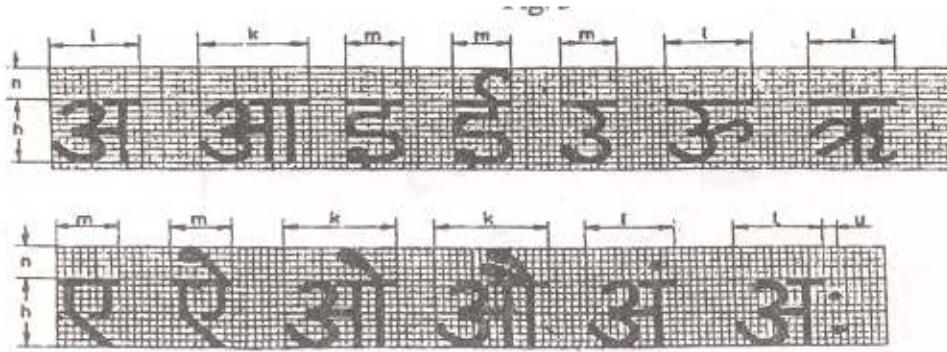
1. <http://www.unicode.org>
2. ISCII'91
3. केंद्रीय हिंदी निदेशालय: देवनागरी लिपि तथा हिंदी वर्तनी का मानकीकरण नई दिल्ली. 2006

## **8. ANNEXURES**

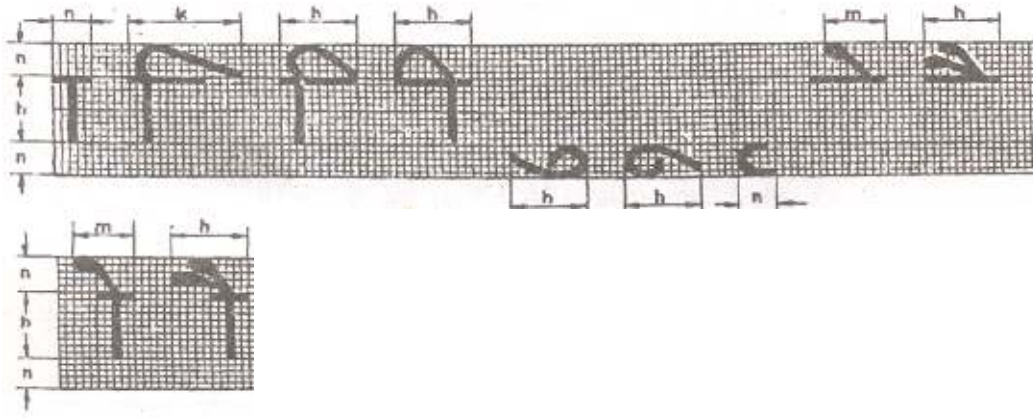
### **Annexure 1: Names of experts who have contributed to the script grammar**



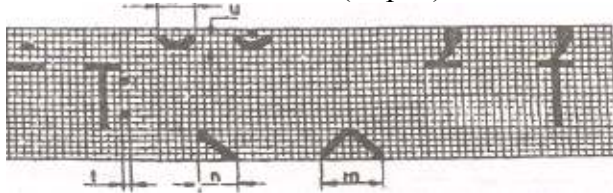
**Annexure 2: Shapes of Hindi characters as per Central Hindi Directorate**  
Hindi Vowels



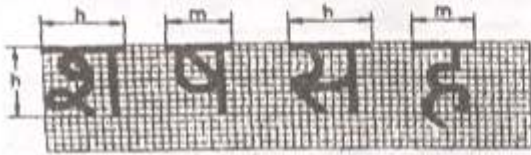
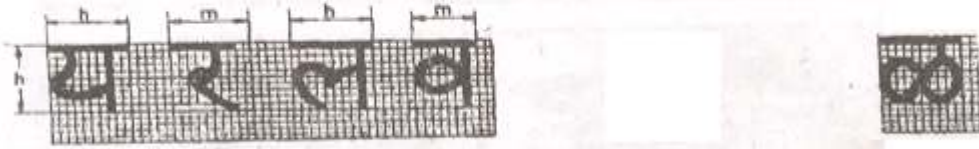
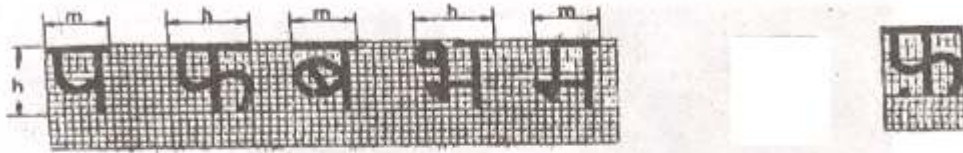
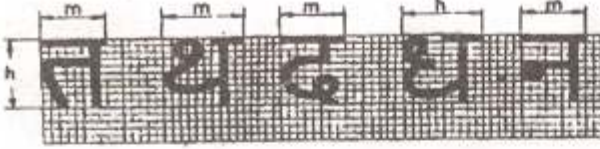
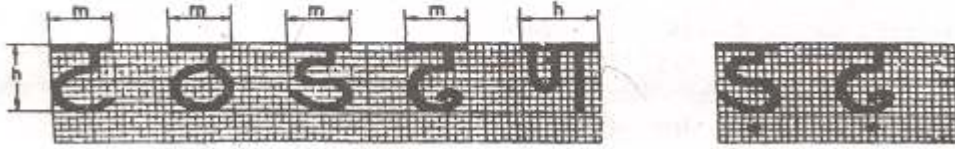
Hindi Vowel Modifiers: Matras



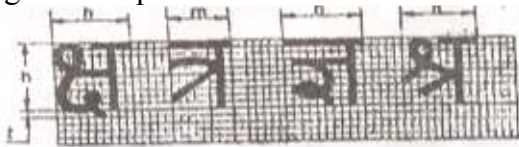
Hindi Diacritics and Rafar(Repha) Rakar



## Hindi Consonants



## Ligatural Representations



### Annexure 3: Unicode Table of Devanāgarī <sup>24</sup>

---

<sup>24</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



