“DEVANAGARI SCRIPT BEHAVIOUR FOR HINDI”

Prepared by

Technology Development for Indian Languages (TDIL) Programme

Department of Electronics and Information Technology, Government of India

in association with

Centre for Development of Advanced Computing (C-DAC)

Table of Contents

0. INTRODUCTION

The term **"Devanagari Script Behaviour for Hindi"** refers to the behaviour pattern of the writing system of Hindi. 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 (linguists, font designers, language experts, academicians) 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 "Devanagari Script behaviour for Hindi" deals with the surface structure of the Hindi 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 that consensually arrive at the "best possible fit" which is acceptable to a majority of users.

Devanagari is a script shared by a large number of languages. In all, apart from Hindi, ten other official languages of India share the same script. However, although these languages share the same matricial script, they differ in the manner in which

a. The choice of a character borrowed from the code-block of Devanagari script.
b. The shape of a given character is represented
c. The ligatural form of a given conjunct is represented
d. The collation order.

Each of these is explained below:

a. Choice of Character:

   Languages differ in the choice of the characters from the Devanagari code-page. Thus Marathi and Konkani use ळ and ऱ (for generating out the eyelash ra). These are not present in Hindi or Dogri. The Hindi ऍ (U+090D) is represented in Marathi and Konkani as ॲ (U+0972). Nukta is used in Hindi and Dogri but not in Marathi or Konkani.

b. The shape of the given character.
   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/ .

   The same is the case with श्र which is represented in Hindi as श्र but in Marathi as श्र. Numbers also display differences

The Hindi number set is as under:

०१२३४५६७८९

The Marathi and Konkani number set is as under:

०१२३४५६७८९

c. The Ligatural shape of the conjunct.
Marathi, Nepali and Konkani prefer stacked shapes of the conjuncts, whereas Hindi as per the directives of the Central Hindi Directorate (q.v.) prefers as far as possible to show conjuncts as linear.
Thus the same ligature is seen as liner if the language is Hindi and as stacked if the language is Nepali or Marathi

Hindi: शक्ति

Nepali/Marathi: शक्ति

d. The collation order within the language.
The collation order[1] varies from language to language although they all share the same script.
In the case of Hindi
क्ष ज्ञ त्र are sorted along with the first consonant of each ligature. Thus क्ष is sorted along with क, ज्ञ with ज and त्र with त

In Marathi क्ष ज्ञ occur at the end of the lexical sort, giving the two conjuncts a specific value of a letter.

In Nepali क्ष ज्ञ त्र are sorted at the end

From the above it will be evident that each language, although it shares the same script, manifests a different behaviour insofar as the implementation of the script is concerned, in terms of its shapes, its ligatural form or even its collation order.

"Devanagari Script behaviour for Hindi" is the term used to define:

- The writing system used to inscribe a Hindi
- The syllabic structure of the writing system of Hindi
- The rule ordering of the characters within the syllable
- Description of the syllabic clusters / ligatures
- Description of valid and invalid clusters i.e. clusters not used within the Hindi
- Collation order of the characters: lexical / dictionary sorting order

As mentioned above Devanāgarī caters to 11 official languages of India. Each of these languages has different representations of the shape of Individual characters as well as their ligatural representations. The "Devanagari Script behaviour for Hindi" in this document is pertinent to Hindi. Other similar documents define the structure of Marathi, Sanskrit, Nepali to name a few languages sharing the common script Devanāgarī and the Unicode code-block 0900-097F (with exception of Rupee Sign and Swastik).

[1] Cf. 6.3.5 of the document

1. OBJECTIVES OF "Devanagari Script behaviour for Hindi"

The objectives of the “Devanagari Script behaviour for Hindi” for 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 left to right order, provide an inventory of displaced catenators, 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. This would serve as a useful input for the font developer, desirous of knowing the visual representation of such combinations and clusters.
- In the case of consonant clusters and ligatural forms, identify to maximum extent such valid combinations and list them out. This would serve as a useful guide for OS developers who would use these to validate or invalidate a combination, entered by the user.
- Provide usage of “Zero Width Joiner (ZWJ)” and “Zero Width Non Joiner (ZWNJ)” within the language. Since ZWJ/ZWNJ are stored in the data, this would ensure that incorrect use of these two signs which could affect search as well as Natural Language Processing, be avoided.

- 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 Marathi 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 "Devanagari Script behaviour for Hindi"

The “Devanagari Script behaviour for Hindi” can be used by a large number of users.

- The primary aim of the “Devanagari Script behaviour for Hindi” has been for the font developer. The “Devanagari Script behaviour for Hindi” 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.

- It allows 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 “Devanagari Script behaviour for Hindi” allows the font designer to write specific rules for such glyphs.

- The other target group is the OS and application developer. Once the possible ligatures and consonant Mātrā combinations have been identified, there is a need to provide a list of maximum combinations within the language.

- Certain features of the “Devanagari Script behaviour for Hindi” such as the shapes can also be used for testing Optical Character Recognition (OCR) and Online Handwriting Recognition (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.

- 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 “Devanagari Script behaviour for Hindi” permits the software developer to write software functions/ routines for sorting data in all applications.

- “Devanagari Script behaviour for Hindi” is equally important for keyboard design, especially when supplemented by frequency data from a corpus.

As can be seen, the “Devanagari Script behaviour for Hindi” 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 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 639-2* & *ISO 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 document.

4. TERMINOLOGY[2]

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.[3] The definition of Abugida can be taken from Unicode chapter 6 instead of Wikipedia

Akshar: see **Abugida.**

Allographs: A variant form of a grapheme that is in complementary distribution or free variation with another form of the same grapheme; an orthographic contextual variant[4]. Thus ae and æ [U+00E6] in Latin alphabet are allographs. Similarly Rafar (repha), Rakar (cf. below) in Indic Scripts are allographs.

AlloScript: The term relates to languages which share a common script. Sub-sets of scripts sharing a single "matricial" script are termed as alloscripts Thus Devanāgarī script is used to write 11 official languages of India. However these languages do not use the same set of characters. Marathi uses the retroflex lla - ळ [U+ 0933] which Hindi does not use. Flaps used in Hindi ड़ [U+095C] ढ़ [U+095D] are not used in Konkani.

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[5]. *A consonant, especially a stop consonant followed by a puff of breath that is clearly audible before the next sound begins.*[6]

Example ख़ in Hindi

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.

---

[2] 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 ISCII91 document and suitably modified where necessary.
[3] Wikipedia definition
[4] http://dictionary.reference.com/browse/allograph
[5] Cf. P 1. 2.1.3. of BIS Document 1991.
[6] http://www.thefreedictionary.com/aspirate. Slightly modified.

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, 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.[7]

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 /ि/ 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 **Eyelash ra).**

Eyelash ra: The eyelash ra or eyebrow ra is an allograph of *ra+Halant* followed by *ya* or *ha* resulting in the following shape of *ra*: र्य र्ह. It is used in Konkani, Nepali and Marathi. In Marathi not all combinations of this type generate an *eyelash ra* e.g. दर्या /darya/ "ocean" vs. दऱ्या /darya/ "valleys". Unicode prescribes a combination of र्+य U+0931 U+094D U+092F र्+ह U+0931 U+094D U+0939 for generating the *eyelash*

[7] Unicode ver. 6.0 Chapter 9.0 pp. 6-7.

ra. Earlier the *eyelash ra* was generated by a combination of ra+halant+ZWJ ⁻ i.e. U+0930 U+094D U+200D.[8] [9]

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**).

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

Nasalized vowel (Anunasika**)**: A vowel pronounced with the flow of air passing both through the nose and the mouth. In Indian scripts this is denoted by a Candrabindu and gives the vowel/vowel sign over which it is placed a nasalized value. Example: जाँच

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

Pure consonant[10]: 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 halant to consonant. The resultant combination places the *ra* on top of the consonant to which it is adjoined e.g. र्+क = र्क In case the consonant itself is adjoined to another consonant, the rafar is placed above the final consonant of the ligature group e.g. र्+घ्+य = र्घ्य. See **Repha.**

Rakar: A special case of a ligature constituted by the adjunction of a consonant followed by a halant to ra. In a large number of Brahmi derived scripts the ra is adjoined to the stem of consonant to which it relates e.g. क्र In the case of consonants which have no

[8] http://unicode.org/~emuller/iwg/p8/utcdoc.html
[9] Cf. p64 of this document
[10] The term used is as per BIS DOCUMENT ***IS 13194: 1991.*** However, it could also be termed as "short obstructive sound" as per suggestion of experts.

stem such as the retroflexes in Devanāgarī, the rakar is placed below the consonant to which it relates e.g. ट्र.

Repha: (see **Rafar).**

Roman script[11]: The script based on the ancient Roman alphabet, with the letters A–Z and a-z (upper and lower case) and also 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: The BIS document defines the vowel as *A letter representing a speech sound made with the vibration of the vocal cords, but without audible obstruction*[12]. In some languages voiceless vowels do occur. Linguistically a Vowel is defined as *a speech sound which is produced by comparatively open configuration of the vocal tract, with vibration of the vocal cords but without audible friction, and which is a unit of the sound system of a language that forms the nucleus of a syllable*[13].

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

[11] The term used is as per BIS DOCUMENT ***IS 13194: 1991.Point No 2.1.20 p. 1***
[12] Cf. P. 1 2.1.6 BIS Document IS 13194. 1991
[13] http://oxforddictionaries.com/definition/english/vowel

5. PHILOSOPHY AND UNDERLYING PRINCIPLES

The "Devanagari Script behaviour for Hindi" is based on the following principles:

1. The document 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 "Devanagari Script behaviour for Hindi" 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[14].

5. The "Devanagari Script behaviour for Hindi" 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.

[14] It is recommended that such variations be culled by placing the document for public review..

6. “Devanagari Script behaviour for Hindi” : STRUCTURE

The “Devanagari Script behaviour for Hindi” 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 “Devanagari Script behaviour for Hindi” proper and describes the character set as well as the conjunct shapes of the given script along with the collation order. Section 6.3.1. which deals with the character set of the language. Sections 6.3.2 and 6.3.3 deal with the Consonant-Mātrā/ Consonant-Mātrā-Nasal combinations and also the 2, 3 and 4 Consonant Ligatures within the language. Combinations of Vowel with Bindi/Anuswara and Candrabindu are also provided.

6.1. PERIPHERAL ELEMENTS OF THE "Devanagari Script behaviour for Hindi"

These constitute the elements that are peripheral to the document. 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 639-2. & 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 क्ल 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 document.

6.2. CONFORMITY TO THE SYLLABLE STRUCTURE[15]

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 it 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	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	ऍ	ऐ	ओ	ऑ	औ

(M) Mātrās

ा	ि	ी	ु	ू	ृ	े	ॅ	ै	ो	ॉ	ौ

(D) Diacritics

ं - Anuswara	Anuswara, an archinasal, is denoted by a dot above the letter after which it is to be pronounced. This falls under Nasal category.
ँ - Candrabindu	Candrabindu is pure nasalization as air comes from the nose. It is denoted by a breve with a dot superposed above

[15] Annexure 5 provides a formalism for the Indic Akshar as represented on the W3CIndia site

	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.
ऽ - Avagraha[16]	For extra length with long vowels as seen in the Sanskrit text /उपदेशेऽजनुनासिक/

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

(N)[17] **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	**C**	N,M,D,H
C	**N**	M,D,H
	V	D
C, N	**M**	D
C, N,V,M	**D**	
C, N	**H**	C

- C can be preceded by H or no subtype and followed by any one of the following: N,M,D,H
- N can be preceded by C and followed by any one of the following: 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,N and followed by D.
- D can be preceded by C, N, V, M and followed by no other subtype. It closes the akshar.
- H can be preceded by C,N 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)

[16] Avagraha is rarely used in Hindi and is used for representing text in Sanskrit where the character is needed..

[17] 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,ख़,ग,ज़,फ़

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.
   1.2. A vowel followed by a modifier, i.e. either an archinasal (*anuswara*) or a visarga: ईं /ĩ /, आः /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[18] (with or without Nukta) followed by a mātrā, i.e. the inherent vowel being substituted by another vowel: की /ki:/,क़ी /qi:/
   2.3. A consonant (with or without Nukta) followed by a modifier: कं /kə̃/, हः /haH[19]/
   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 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 *sanyuktakshara.*

        त्क /tka/ त्कं /tkə̃/, त्कः /tkaH/ त्कुं /tkũ/, 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: 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[20]	CM	CD	CMD
CHC	CHCM	CHCD	CHCMD
CHCHC	CHCHCM	CHCHCD	CHCHCMD

[18] For purposes of Simplification, C here will automatically be treated as being also consonant+nukta: C+ N

[19] This character represents phonetically the weak implicit vowel, termed as schwa and often shown as /a/ also.

[20] C here will automatically be treated as being also consonant+nukta, C+N to simplify the explanation

CHCHCHC	CHCHCHCM	CHCHCHCD	CHCHCHCM

A total number of 16 theoretical syllables is therefore possible.

Since the formal structure “Devanagari Script behaviour for Hindi” 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.

6.3 "DEVANAGARI SCRIPT BEHAVIOUR FOR HINDI" PROPER

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

6.3.1. The Character Set of Hindi.
6.3.2. The Consonant mātrā combinations of Hindi as well as Vowel and nasal modifier combinations.
6.3.3. The Ligature Set of Hindi.
6.3.4 Inventory of Valid and Invalid Combinations with respect to 6.3.2. and 6.3.3.
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 "Devanagari Script behaviour for Hindi".
This comprises the following:

6.3.1.1. The Consonant Set.[21]
6.3.1.2. The Vowel Set.
6.3.1.3. The Mātrā Set.
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. Halant
6.3.1.8. Numerals.
6.3.1.9. Punctuation marks.
6.3.1.10. Other symbols.

[21] The shapes provided here are as desired by Central Hindi Directorate. These are provided in Appendix 1

Each of these will be analyzed 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[22].

	STOPS				NASAL	TAP	LATERAL	FRICATIVE		APPROXIMANT
	-vd -asp	-vd +asp	+vd -asp	+vd +asp	+vd -asp	+vd	+vd	-vd	+vd	
Laryngeal								ह		
Uvular	क़		ग़							
Velar	क	ख	ग	घ	ङ			ख़		
Palato-alveolar								श		
Palatal	च	छ	ज	झ	ञ					य
Retro-flex	ट	ठ	ड	ढ	ण			ष		
Flaps			ड़	ढ़						
Dental/ Alveolar[23]	त	थ	द	ध	न	र	ल	स	ज़	
Labio-Dental								फ़		व
Bi-labial	प	फ	ब	भ	म					

Note: Ligatures क्ष त्र ज्ञ श्र are not listed in the consonants list[24].

[22] Based on comments and suggestions by Dr Pramod Pandey Professor, Centre for Linguistics, School of Languages, Literature and Culture Studies, Jawaharlal Nehru University to make the chart representative of Modern Hindi..

[23] The two points of articulation are placed together for sake of economy.

[24] These are treated as Ligatures by the Central Hindi Directorate cf. p 2 and also p.5 of their document

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

	STOPS				NASAL	TAP	LATERAL	FRICATIVE		APPROXIMANT
	-vd -asp	-vd +asp	+vd -asp	+vd +asp	+vd -asp	+vd	+vd	-vd	+vd	
Laryngeal								ह		
Uvular	क़		ग़							
Velar	क	ख	ग	घ	ङ			ख़		
Palato-alveolar								श		
Palatal	च	छ	ज	झ	ञ					य
Retro-flex	ट	ठ	ड	ढ	ण			ष		
Flaps			ड़	ढ़						
Dental/ Alveolar	त	थ	द	ध	न	र	ल	स	ज़	
Labio-Dental								फ़		व
Bi-labial	प	फ	ब	भ	म					

6.3.1.2. The Vowel Set

The Vowel set of Hindi is as under:

Character	Unicode code-point	Character name
अ	U+0905	DEVANĀGARĪ LETTER A
आ	U+0906	DEVANĀGARĪ LETTER AA
इ	U+0907	DEVANĀGARĪ LETTER I
ई	U+0908	DEVANĀGARĪ LETTER II
उ	U+0909	DEVANĀGARĪ LETTER U
ऊ	U+090A	DEVANĀGARĪ LETTER UU
ऋ	U+090B	DEVANĀGARĪ LETTER VOCALIC R
ए	U+090F	DEVANĀGARĪ LETTER E
ऍ	U+090D	DEVANĀGARĪ LETTER CANDRA E
ऐ	U+0910	DEVANĀGARĪ LETTER AI
ओ	U+0913	DEVANĀGARĪ LETTER O
ऑ	U+0911	DEVANĀGARĪ LETTER CANDRA O
औ	U+0914	DEVANĀGARĪ LETTER AU

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

अ आ इ ई उ ऊ ऋ ए ऐ ऍ ओ ऑ औ

6.3.1.3. The Mātrā Set

The Mātrā[25] (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 Candra E	◌ॅ	ऍ	क् +◌ॅ = कॅ
9. Devanāgarī sign AI	◌ै	ऐ	क् + ऐ = कै
10. Devanāgarī sign O	◌ो	ओ	क् + ओ = को
11. Devanāgarī sign Candra O	◌ॉ	ऑ	क् +◌ॉ = कॉ
12. Devanāgarī sign AU	◌ौ	औ	क् + औ = कौ

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

◌ा ◌ि ◌ी ◌ु ◌ू ◌ृ ◌े ◌ॅ ◌ै ◌ो ◌ॉ ◌ौ

[25] 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 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: र्क र्त e.g. top rafars will be formed in case of following words.

धर्म, चर्ख़ा

RAKARS

1. Bottom rakar ट्र ड्र
2. Side rakar क्र च्र
3. Inside rakar ह्न

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

Bottom rakar	ड्रम, राष्ट्र
Side rakar	व्रत, चक्र, प्रसाद,स्रोत
Inside rakar	ह्रस्व, ह्रास

6.3.1.6. Diacritics

These are as under in the case of Hindi:

ं : - Anuswar रंग

ँ : - Candrabindu/Anunasika इँट

ः - Visarga दुःख

ऽ: - For extra length with long vowels e.g. / उपदेशेऽजनुनासिक / [26]

6.3.1.7. Halant

् : - Halant संवत्

6.3.1.8. 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.

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

[26] Avagraha is rarely used in Hindi and is used for representing text in Sanskrit here the character is needed.. The example cited is from http://hi.wikipedia.org/s/193x

6.3.1.9. Punctuation Markers

- Hindi uses punctuation markers from the Latin set. such as . , ; : " ' ( ) [ ] etc.
- However, the abbreviation marker(॰ U+0970) is often used in Devanāgarī
- Purna and Deergha Virama (full-stop/danda) Devanāgarī code block: U+0964, U+0965 ।,॥. However, Unicode recognizes these characters as DEVANAGARI DANDA (।) and DEVANAGARI DOUBLE DANDA (॥).
- । is used to mark the full stop and "। and ॥ are used for poetry written in *Chhanda/Doha/Chaupai*" . The usage of the same is shown in the Doha of Sant Kabir given below

चाह मिटी, चिंता मिटी मनवा बेपरवाह । जिसको कुछ नहीं चाहिए वह शहंशाह ॥
माटी कहे कुम्हार से, तू क्या रौंदे मोय । इक दिन ऐसा आएगा, मैं रौंदूँगी 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 Purna Viram	।
21.	Devanāgarī Deergha Viram	।।

6.3.1.10 Other Symbols

These are religious symbols and currency symbol included in Unicode:

ॐ: Om (as written in Hindi) (Unicode code point: 0950)

₹: Rupee Sign as mandated by Government of India. (Unicode code point: 20B9)

卐: Right-facing svasti sign (Unicode code point: 0FD5)

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:	क़	ख़	ग़	ज़	फ़	ड़	ढ़			

All valid as well as invalid combinations have been provided since the primary aim of the document is to ensure that the font developer can develop a valid font based on the combinations. The "Devanagari Script Behaviour for Hindi" 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.

It needs to be noted that the font developer requires to represent within the Open Type Table rules for all characters and even if the combinations of certain characters do not exist in the language, such combinations need to be represented to enable font developers to show the exact representation of these combinations. Such "dead" combinations may not be valid within the language system but are a precious resource for the font developer and hence have been included.

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: ङा

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 Candrabindu. 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	ञा
ि	कि	खि	गि	घि	ङि	चि	छि	जि	झि	ञि
ी	की	खी	गी	घी	ङी	ची	छी	जी	झी	ञी
ु	कु	खु	गु	घु	ङु	चु	छु	जु	झु	ञु
ू	कू	खू	गू	घू	ङू	चू	छू	जू	झू	ञू
ृ	कृ	खृ	गृ	घृ	ङृ	चृ	छृ	जृ	झृ	ञृ
े	के	खे	गे	घे	ङे	चे	छे	जे	झे	ञे
ॅ	कॅ	खॅ	गॅ	घॅ	ङॅ	चॅ	छॅ	जॅ	झॅ	ञॅ
ै	कै	खै	गै	घै	ङै	चै	छै	जै	झै	ञै
ो	को	खो	गो	घो	ङो	चो	छो	जो	झो	ञो
ॉ	कॉ	खॉ	गॉ	घॉ	ङॉ	चॉ	छॉ	जॉ	झॉ	ञॉ
ौ	कौ	खौ	गौ	घौ	ङौ	चौ	छौ	जौ	झौ	ञौ

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 mātrā 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
ॅ	टॅ	ठॅ	डॅ	ढॅ	णॅ	तॅ	थॅ	दॅ	धॅ	नॅ
ै	टै	ठै	डै	ढै	णै	तै	थै	दै	धै	नै
ो	टो	ठो	डो	ढो	णो	तो	थो	दो	धो	नो
ॉ	टॉ	ठॉ	डॉ	ढॉ	णॉ	तॉ	थॉ	दॉ	धॉ	नॉ
ौ	टौ	ठौ	डौ	ढौ	णौ	तौ	थौ	दौ	धौ	नौ

Consonant and Mātrā combinations Set 3

This set is in continuation of set 2 which shows consonant and mātrā 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	यृ	रृ	लृ	वृ	शृ	षृ	सृ	हृ
◌े	पे	फे	बे	भे	मे	ये	रे	ले	वे	शे	षे	से	हे
◌ॅ	पॅ	फॅ	बॅ	भॅ	मॅ	यॅ	रॅ	लॅ	वॅ	शॅ	षॅ	सॅ	हॅ
◌ै	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
◌ौ	पौ	फौ	बौ	भौ	मौ	यौ	रौ	लौ	वौ	शौ	षौ	सौ	हौ

Consonant and Mātrā combinations Set 4

This set is in continuation of set 3 which shows consonant and mātrā combinations for nukta consonants.[27]

	क़	ख़	ग़	ज़	फ़	ड़	ढ़
ा	क़ा	ख़ा	ग़ा	ज़ा	फ़ा	ड़ा	ढ़ा
ि	क़ि	ख़ि	ग़ि	ज़ि	फ़ि	ड़ि	ढ़ि
ी	क़ी	ख़ी	ग़ी	ज़ी	फ़ी	ड़ी	ढ़ी
ु	क़ु	ख़ु	ग़ु	ज़ु	फ़ु	ड़ु	ढ़ु
ू	क़ू	ख़ू	ग़ू	ज़ू	फ़ू	ड़ू	ढ़ू
ृ[28]	क़ृ	ख़ृ	ग़ृ	ज़ृ	फ़ृ	ड़ृ	ढ़ृ
े	क़े	ख़े	ग़े	ज़े	फ़े	ड़े	ढ़े
ॅ	क़ॅ	ख़ॅ	ग़ॅ	ज़ॅ	फ़ॅ	ड़ॅ	ढ़ॅ
ै	क़ै	ख़ै	ग़ै	ज़ै	फ़ै	ड़ै	ढ़ै
ो	क़ो	ख़ो	ग़ो	ज़ो	फ़ो	ड़ो	ढ़ो
ॉ	क़ॉ	ख़ॉ	ग़ॉ	ज़ॉ	फ़ॉ	ड़ॉ	ढ़ॉ
ौ	क़ौ	ख़ौ	ग़ौ	ज़ौ	फ़ौ	ड़ौ	ढ़ौ

[27] क़,ग़ were not recognized as part of the consonant set by the Central Hindi Directorate, but have been accepted in its latest version..

[28] 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	ञां
◌िं	किं	खिं	गिं	घिं	ङिं	चिं	छिं	जिं	झिं	ञिं
◌ीं	कीं	खीं	गीं	घीं	ङीं	चीं	छीं	जीं	झीं	ञीं
◌ुं	कुं	खुं	गुं	घुं	ङुं	चुं	छुं	जुं	झुं	ञुं
◌ूं	कूं	खूं	गूं	घूं	ङूं	चूं	छूं	जूं	झूं	ञूं
◌ृं	कृं	खृं	गृं	घृं	ङृं	चृं	छृं	जृं	झृं	ञृं
◌ें	कें	खें	गें	घें	ङें	चें	छें	जें	झें	ञें
◌ॅं	कॅं	खॅं	गॅं	घॅं	ङॅं	चॅं	छॅं	जॅं	झॅं	ञॅं
◌ैं	कैं	खैं	गैं	घैं	ङैं	चैं	छैं	जैं	झैं	ञैं
◌ों	कों	खों	गों	घों	ङों	चों	छों	जों	झों	ञों
◌ॉं	कॉं	खॉं	गॉं	घॉं	ङॉं	चॉं	छॉं	जॉं	झॉं	ञॉं
◌ौं	कौं	खौं	गौं	घौं	ङौं	चौं	छौं	जौं	झौं	ञौं

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
◌ृं	टृं	ठृं	डृं	ढृं	णृं	तृं	थृं	दृं	धृं	नृं
◌ें	टें	ठें	डें	ढें	णें	तें	थें	दें	धें	नें
◌ॅं[29]	टॅं	ठॅं	डॅं	ढॅं	णॅं	तॅं	थॅं	दॅं	धॅं	नॅं
◌ैं	टैं	ठैं	डैं	ढैं	णैं	तैं	थैं	दैं	धैं	नैं
◌ों	टों	ठों	डों	ढों	णों	तों	थों	दों	धों	नों
◌ॉं	टॉं	ठॉं	डॉं	ढॉं	णॉं	तॉं	थॉं	दॉं	धॉं	नॉं
◌ौं	टौं	ठौं	डौं	ढौं	णौं	तौं	थौं	दौं	धौं	नौं

[29] ॅ,◌ॅ 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
◌ू◌ं	पूं	फूं	बूं	भूं	मूं	यूं	रूं	लूं	वूं	शूं	षूं	सूं	हूं
◌ृ◌ं	पृं	फृं	बृं	भृं	मृं	यृं	रृं	लृं	वृं	शृं	षृं	सृं	हृं
◌े◌ं	पें	फें	बें	भें	में	यें	रें	लें	वें	शें	षें	सें	हें
◌ॅ◌ं[30]	पॅं	फॅं	बॅं	भॅं	मॅं	यॅं	रॅं	लॅं	वॅं	शॅं	षॅं	सॅं	हॅं
◌ै◌ं	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
◌ॉ◌ं	पॉं	फॉं	बॉं	भॉं	मॉं	यॉं	रॉं	लॉं	वॉं	शॉं	षॉं	सॉं	हॉं
◌ौ◌ं	पौं	फौं	बौं	भौं	मौं	यौं	रौं	लौं	वौं	शौं	षौं	सौं	हौं

[30] 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	ख़	ग	ज़	फ़	ड़	ढ़
◌ं	क़ं	ख़ं	ग़ं	ज़ं	फ़ं	ड़ं	ढ़ं
◌ा	क़ा	ख़ा	ग़ा	ज़ा	फ़ा	ड़ा	ढ़ा
◌िं	क़िं	ख़िं	ग़िं	ज़िं	फ़िं	ड़िं	ढ़िं
◌ीं	क़ीं	ख़ीं	ग़ीं	ज़ीं	फ़ीं	ड़ीं	ढ़ीं
◌ुं	क़ुं	ख़ुं	ग़ुं	ज़ुं	फ़ुं	ड़ुं	ढ़ुं
◌ूं	क़ूं	ख़ूं	ग़ूं	ज़ूं	फ़ूं	ड़ूं	ढ़ूं
◌ृं	क़ृं	ख़ृं	ग़ृं	ज़ृं	फ़ृं	ड़ृं	ढ़ृं
◌ें	क़ें	ख़ें	ग़ें	ज़ें	फ़ें	ड़ें	ढ़ें
◌ॅं[31]	क़ॅं	ख़ॅं	ग़ॅं	ज़ॅं	फ़ॅं	ड़ॅं	ढ़ॅं
◌ैं	क़ैं	ख़ैं	ग़ैं	ज़ैं	फ़ैं	ड़ैं	ढ़ैं
◌ों	क़ों	ख़ों	ग़ों	ज़ों	फ़ों	ड़ों	ढ़ों
◌ॉं	क़ॉं	ख़ॉं	ग़ॉं	ज़ॉं	फ़ॉं	ड़ॉं	ढ़ॉं
◌ौं	क़ौं	ख़ौं	ग़ौं	ज़ौं	फ़ौं	ड़ौं	ढ़ौं

[31] 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 Candrabindu[32] - 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	ञाँ
िँ	किं	खिं	गिं	घिं	ङिं	चिं	छिं	जिं	झिं	ञिं
ीँ	कीं	खीं	गीं	घीं	ङीं	चीं	छीं	जीं	झीं	ञीं
ुँ	कुँ	खुँ	गुँ	घुँ	ङुँ	चुँ	छुँ	जुँ	झुँ	ञुँ
ूँ	कूं	खूं	गूं	घूं	ङूं	चूं	छूं	जूं	झूं	ञूं
ृँ	कृँ	खृँ	गृँ	घृँ	ङृँ	चृँ	छृँ	जृँ	झृँ	ञृँ
ेँ	कें	खें	गें	घें	ङें	चें	छें	जें	झें	ञें
ॅँ	कॅं	खॅं	गॅं	घॅं	ङॅं	चॅं	छॅं	जॅं	झॅं	ञॅं
ैँ	कैं	खैं	गैं	घैं	ङैं	चैं	छैं	जैं	झैं	ञैं
ोँ	कों	खों	गों	घों	ङों	चों	छों	जों	झों	ञों
ॉँ	कॉं	खॉं	गॉं	घॉं	ङॉं	चॉं	छॉं	जॉं	झॉं	ञॉं
ौँ	कौं	खौं	गौं	घौं	ङौं	चौं	छौं	जौं	झौं	ञौं

[32] As per rule 2.6.2.3. of the Central Hindi Directorate, Candrabindu cannot be placed over matras which are above the Shirorekha and in this case the Candrabindu is replaced by an Anuswar.

Consonant and Mātrā +Nasal combinations With Candrabindu - Set 2
This set is in continuation of set 1 above which shows combinations of Consonant and Mātrā + Candrabindu.

	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
िँ	टिं	ठिं	डिं	ढिं	णिं	तिं	थिं	दिं	धिं	निं
ीँ	टीं	ठीं	डीं	ढीं	णीं	तीं	थीं	दीं	धीं	नीं
ुँ	टुँ	ठुँ	डुँ	ढुँ	णुँ	तुँ	थुँ	दुँ	धुँ	नुँ
ूँ	टूँ	ठूँ	डूँ	ढूँ	णूँ	तूँ	थूँ	दूँ	धूँ	नूँ
ृँ	टृँ	ठृँ	डृँ	ढृँ	णृँ	तृँ	थृँ	दृँ	धृँ	नृँ
ेँ	टें	ठें	डें	ढें	णें	तें	थें	दें	धें	नें
ॅँ	टॅं	ठॅं	डॅं	ढॅं	णॅं	तॅं	थॅं	दॅं	धॅं	नॅं
ैँ	टैं	ठैं	डैं	ढैं	णैं	तैं	थैं	दैं	धैं	नैं
ोँ	टों	ठों	डों	ढों	णों	तों	थों	दों	धों	नों
ॉँ[33]	टॉं	ठॉं	डॉं	ढॉं	णॉं	तॉं	थॉं	दॉं	धॉं	नॉं
ौँ	टौं	ठौं	डौं	ढौं	णौं	तौं	थौं	दौं	धौं	नौं

[33] 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 Candrabindu - Set 3
This set is in continuation of set 2 above which shows combinations of Consonant and Mātrā + Candrabindu.

	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
िँ	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
ुँ	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	ऱृँ	लृँ	वृँ	शृँ	षृँ	सृँ	ह्रँ
ेँ	पें	फें	बें	भें	में	यें	रें	लें	वें	शें	षें	सें	हें
ॅँ	पॅं	फॅं	बॅं	भॅं	मॅं	यॅं	रॅं	लॅं	वॅं	शॅं	षॅं	सॅं	हॅं
ैँ	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
ॉँ	पॉं	फॉं	बॉं	भॉं	मॉं	यॉं	रॉं	लॉं	वॉं	शॉं	षॉं	सॉं	हॉं
ौँ	पौं	फौं	बौं	भौं	मौं	यौं	रौं	लौं	वौं	शौं	षौं	सौं	हौं

Consonant and Mātrā +Nasal combinations With Candrabindu - Set 4

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

	क़	ख़	ग़	ज़	फ़	ड़	ढ़
ँ	क़ँ	ख़ँ	ग़ँ	ज़ँ	फ़ँ	ड़ँ	ढ़ँ
ाँ	क़ाँ	ख़ाँ	ग़ाँ	ज़ाँ	फ़ाँ	ड़ाँ	ढ़ाँ
िँ	क़िं	ख़िं	ग़िं	ज़िं	फ़िं	ड़िं	ढ़िं
ीँ	क़ीं	ख़ीं	ग़ीं	ज़ीं	फ़ीं	ड़ीं	ढ़ीं
ुँ	क़ुँ	ख़ुँ	ग़ुँ	ज़ुँ	फ़ुँ	ड़ुँ	ढ़ुँ
ूँ	क़ूँ	ख़ूँ	ग़ूँ	ज़ूँ	फ़ूँ	ड़ूँ	ढ़ूँ
ृँ	क़ृँ	ख़ृँ	ग़ृँ	ज़ृँ	फ़ृँ	ड़ृँ	ढ़ृँ
ेँ	क़ें	ख़ें	ग़ें	ज़ें	फ़ें	ड़ें	ढ़ें
ॅँ	क़ॅं	ख़ॅं	ग़ॅं	ज़ॅं	फ़ॅं	ड़ॅं	ढ़ॅं
ैँ	क़ैं	ख़ैं	ग़ैं	ज़ैं	फ़ैं	ड़ैं	ढ़ैं
ोँ	क़ों	ख़ों	ग़ों	ज़ों	फ़ों	ड़ों	ढ़ों
ॉँ	क़ॉं	ख़ॉं	ग़ॉं	ज़ॉं	फ़ॉं	ड़ॉं	ढ़ॉं
ौँ	क़ौं	ख़ौं	ग़ौं	ज़ौं	फ़ौं	ड़ौं	ढ़ौं

Vowel+Nasal Combinations: Anuswar and Candrabindu

The table below shows the combinations of full vowels with the nasal modifiers: Anuswar and Candrabindu.

	अ	आ	इ	ई	उ	ऊ	ऋ	ऍ	ए	ऐ	ऑ	ओ	औ
ँ	अँ	आँ	इँ	ईँ	उँ	ऊँ	ऋँ	ऍं	एँ	ऐं	ऑं	ओं	औं
ं	अं	आं	इं	ईं	उं	ऊं	ऋं	ऍं	एं	ऐं	ऑं	ओं	औं

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.

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 e.g. Combination of consonant "क" and " क" joined by a Halant is the ligature "क्क[34]".

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	ञ्ञ
च्	च्क	च्ख	च्ग	च्घ	च्ङ	च्च	च्छ	च्ज	च्झ	च्ञ
छ्	छ्क	छ्ख	छ्ग	छ्घ	छ्ङ	छ्च	छ्छ	छ्ज	छ्झ	छ्ञ
ज्	ज्क	ज्ख	ज्ग	ज्घ	ज्ङ	ज्च	ज्छ	ज्ज	ज्झ	ज़
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

[34] ZWJ/ZWNJ are generally not used in Hindi. cf. Annexure 3: Note on ZWJ/ZWNJ in Devanāgarī.

ड़्	ड़्क	ड़्ख	ड़्ग	ड़्घ	ड़्ङ	ड़्च	ड़्छ	ड़्ज	ड़्झ	ड़्ञ
ढ़्	ढ़्क	ढ़्ख	ढ़्ग	ढ़्घ	ढ़्ङ	ढ़्च	ढ़्छ	ढ़्ज	ढ़्झ	ढ़्ञ
ण्	ण्क	ण्ख	ण्ग	ण्घ	ण्ङ	ण्च	ण्छ	ण्ज	ण्झ	ण्ञ
त्	त्क	त्ख	त्ग	त्घ	त्ङ	त्च	त्छ	त्ज	त्झ	त्ञ
थ्	थ्क	थ्ख	थ्ग	थ्घ	थ्ङ	थ्च	थ्छ	थ्ज	थ्झ	थ्ञ
द्	द्क	द्ख	द्ग	द्घ	द्ङ	द्च	द्छ	द्ज	द्झ	द्ञ
ध्	ध्क	ध्ख	ध्ग	ध्घ	ध्ङ	ध्च	ध्छ	ध्ज	ध्झ	ध्ञ
न्	न्क	न्ख	न्ग	न्घ	न्ङ	न्च	न्छ	न्ज	न्झ	न्ञ
प्	प्क	प्ख	प्ग	प्घ	प्ङ	प्च	प्छ	प्ज	प्झ	प्ञ
फ्	फ्क	फ्ख	फ्ग	फ्घ	फ्ङ	फ्च	फ्छ	फ्ज	फ्झ	फ्ञ
ब्	ब्क	ब्ख	ब्ग	ब्घ	ब्ङ	ब्च	ब्छ	ब्ज	ब्झ	ब्ञ
भ्	भ्क	भ्ख	भ्ग	भ्घ	भ्ङ	भ्च	भ्छ	भ्ज	भ्झ	भ्ञ
म्	म्क	म्ख	म्ग	म्घ	म्ङ	म्च	म्छ	म्ज	म्झ	म्ञ
य्	य्क	य्ख	य्ग	य्घ	य्ङ	य्च	य्छ	य्ज	य्झ	य्ञ
र्	र्क	र्ख	र्ग	र्घ	र्ङ	र्च	र्छ	र्ज	र्झ	र्ञ
ल्	ल्क	ल्ख	ल्ग	ल्घ	ल्ङ	ल्च	ल्छ	ल्ज	ल्झ	ल्ञ
व्	व्क	व्ख	व्ग	व्घ	व्ङ	व्च	व्छ	व्ज	व्झ	व्ञ
श्	श्क	श्ख	श्ग	श्घ	श्ङ	श्च	श्छ	श्ज	श्झ	श्ञ
ष्	ष्क	ष्ख	ष्ग	ष्घ	ष्ङ	ष्च	ष्छ	ष्ज	ष्झ	ष्ञ
स्	स्क	स्ख	स्ग	स्घ	स्ङ	स्च	स्छ	स्ज	स्झ	स्ञ
ह्	ह्क	ह्ख	ह्ग	ह्घ	ह्ङ	ह्च	ह्छ	ह्ज	ह्झ	ह्ञ
क़्	क़्क	क़्ख	क़्ग	क़्घ	क़्ङ	क़्च	क़्छ	क़्ज	क़्झ	क़्ञ
ख़्	ख़्क	ख़्ख	ख़्ग	ख़्घ	ख़्ङ	ख़्च	ख़्छ	ख़्ज	ख़्झ	ख़्ञ
ग़्	ग़्क	ग़्ख	ग़्ग	ग़्घ	ग़्ङ	ग़्च	ग़्छ	ग़्ज	ग़्झ	ग़्ञ
ज़्	ज़्क	ज़्ख	ज़्ग	ज़्घ	ज़्ङ	ज़्च	ज़्छ	ज़्ज	ज़्झ	ज़्ञ
फ़्	फ़्क	फ़्ख	फ़्ग	फ़्घ	फ़्ङ	फ़्च	फ़्छ	फ़्ज	फ़्झ	फ़्ञ
ड़्	ड़्क	ड़्ख	ड़्ग	ड़्घ	ड़्ङ	ड़्च	ड़्छ	ड़्ज	ड़्झ	ड़्ञ
ढ़्	ढ़्क	ढ़्ख	ढ़्ग	ढ़्घ	ढ़्ङ	ढ़्च	ढ़्छ	ढ़्ज	ढ़्झ	ढ़्ञ

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 e.g. Combination of consonant "क" and " ट" is ligature "क्ट".

CHC(combination of two consonants) - Set 2										
	ट	ठ	ड	ढ	ण	त	थ	द	ध	न
क्	क्ट	क्ठ	क्ड	क्ढ	क्ण	क्त	क्थ	क्द	क्ध	क्न
ख्	ख्ट	ख्ठ	ख्ड	ख्ढ	ख्ण	ख्त	ख्थ	ख्द	ख्ध	ख्न
ग्	ग्ट	ग्ठ	ग्ड	ग्ढ	ग्ण	ग्त	ग्थ	ग्द	ग्ध	ग्न
घ्	घ्ट	घ्ठ	घ्ड	घ्ढ	घ्ण	घ्त	घ्थ	घ्द	घ्ध	घ्न
ङ्	ङ्ट	ङ्ठ	ङ्ड	ङ्ढ	ङ्ण	ङ्त	ङ्थ	ङ्द	ङ्ध	ङ्न
च्	च्ट	च्ठ	च्ड	च्ढ	च्ण	च्त	च्थ	च्द	च्ध	च्न
छ्	छ्ट	छ्ठ	छ्ड	छ्ढ	छ्ण	छ्त	छ्थ	छ्द	छ्ध	छ्न
ज्	ज्ट	ज्ठ	ज्ड	ज्ढ	ज्ण	ज्त	ज्थ	ज्द	ज्ध	ज्न
झ्	झ्ट	झ्ठ	झ्ड	झ्ढ	झ्ण	झ्त	झ्थ	झ्द	झ्ध	झ्न
ञ्	ञ्ट	ञ्ठ	ञ्ड	ञ्ढ	ञ्ण	ञ्त	ञ्थ	ञ्द	ञ्ध	ञ्न
ट्	ट्ट	ट्ठ	ट्ड	ट्ढ	ट्ण	ट्त	ट्थ	ट्द	ट्ध	ट्न
ठ्	ठ्ट	ठ्ठ	ठ्ड	ठ्ढ	ठ्ण	ठ्त	ठ्थ	ठ्द	ठ्ध	ठ्न
ड्	ड्ट	ड्ठ	ड्ड	ड्ढ	ड्ण	ड्त	ड्थ	ड्द	ड्ध	ड्न
ढ्	ढ्ट	ढ्ठ	ढ्ड	ढ्ढ	ढ्ण	ढ्त	ढ्थ	ढ्द	ढ्ध	ढ्न
ण्	ण्ट	ण्ठ	ण्ड	ण्ढ	ण्ण	ण्त	ण्थ	ण्द	ण्ध	ण्न
त्	त्ट	त्ठ	त्ड	त्ढ	त्ण	त्त	त्थ	त्द	त्ध	त्न
थ्	थ्ट	थ्ठ	थ्ड	थ्ढ	थ्ण	थ्त	थ्थ	थ्द	थ्ध	थ्न
द्	द्ट	द्ठ	द्ड	द्ढ	द्ण	द्त	द्थ	द्द	द्ध	द्न
ध्	ध्ट	ध्ठ	ध्ड	ध्ढ	ध्ण	ध्त	ध्थ	ध्द	ध्ध	ध्न
न्	न्ट	न्ठ	न्ड	न्ढ	न्ण	न्त	न्थ	न्द	न्ध	न्न
प्	प्ट	प्ठ	प्ड	प्ढ	प्ण	प्त	प्थ	प्द	प्ध	प्न
फ्	फ्ट	फ्ठ	फ्ड	फ्ढ	फ्ण	फ्त	फ्थ	फ्द	फ्ध	फ्न

ब्	ब्ट	ब्ठ	ब्ड	ब्ढ	ब्ण	ब्त	ब्थ	ब्द	ब्ध	ब्न
भ्	भ्ट	भ्ठ	भ्ड	भ्ढ	भ्ण	भ्त	भ्थ	भ्द	भ्ध	भ्न
म्	म्ट	म्ठ	म्ड	म्ढ	म्ण	म्त	म्थ	म्द	म्ध	म्न
य्	य्ट	य्ठ	य्ड	य्ढ	य्ण	य्त	य्थ	य्द	य्ध	य्न
र्	र्ट	र्ठ	र्ड	र्ढ	र्ण	र्त	र्थ	र्द	र्ध	र्न
ल्	ल्ट	ल्ठ	ल्ड	ल्ढ	ल्ण	ल्त	ल्थ	ल्द	ल्ध	ल्न
व्	व्ट	व्ठ	व्ड	व्ढ	व्ण	व्त	व्थ	व्द	व्ध	व्न
श्	श्ट	श्ठ	श्ड	श्ढ	श्ण	श्त	श्थ	श्द	श्ध	श्न
ष्	ष्ट	ष्ठ	ष्ड	ष्ढ	ष्ण	ष्त	ष्थ	ष्द	ष्ध	ष्न
स्	स्ट	स्ठ	स्ड	स्ढ	स्ण	स्त	स्थ	स्द	स्ध	स्न
ह्	ह्ट	ह्ठ	ह्ड	ह्ढ	ह्ण	ह्त	ह्थ	ह्द	ह्ध	ह्न
क़्	क़्ट	क़्ठ	क़्ड	क़्ढ	क़्ण	क़्त	क़्थ	क़्द	क़्ध	क़्न
ख़्	ख़्ट	ख़्ठ	ख़्ड	ख़्ढ	ख़्ण	ख़्त	ख़्थ	ख़्द	ख़्ध	ख़्न
ग़्	ग़्ट	ग़्ठ	ग़्ड	ग़्ढ	ग़्ण	ग़्त	ग़्थ	ग़्द	ग़्ध	ग़्न
ज़्	ज़्ट	ज़्ठ	ज़्ड	ज़्ढ	ज़्ण	ज़्त	ज़्थ	ज़्द	ज़्ध	ज़्न
फ़्	फ़्ट	फ़्ठ	फ़्ड	फ़्ढ	फ़्ण	फ़्त	फ़्थ	फ़्द	फ़्ध	फ़्न
ड़्	ड़्ट	ड़्ठ	ड़्ड	ड़्ढ	ड़्ण	ड़्त	ड़्थ	ड़्द	ड़्ध	ड़्न
ढ़्	ढ़्ट	ढ़्ठ	ढ़्ड	ढ़्ढ	ढ़्ण	ढ़्त	ढ़्थ	ढ़्द	ढ़्ध	ढ़्न

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 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	ख्प	ख्फ	ख्ब	ख्भ	ख्म	ख्य	ख्र	ख्ल	ख्व	ख्श	ख्ष	ख्स	ख्ह
ग्	ग्प	ग्फ	ग्ब	ग्भ	ग्म	ग्य	ग्र	ग्ल	ग्व	ग्श	ग्ष	ग्स	ग्ह
घ्	घ्प	घ्फ	घ्ब	घ्भ	घ्म	घ्य	घ्र	घ्ल	घ्व	घ्श	घ्ष	घ्स	घ्ह
ङ्	ङ्प	ङ्फ	ङ्ब	ङ्भ	ङ्म	ङ्य	ङ्र	ङ्ल	ङ्व	ङ्श	ङ्ष	ङ्स	ङ्ह
च्	च्प	च्फ	च्ब	च्भ	च्म	च्य	च्र	च्ल	च्व	च्श	च्ष	च्स	च्ह
छ्	छ्प	छ्फ	छ्ब	छ्भ	छ्म	छ्य	छ्र	छ्ल	छ्व	छ्श	छ्ष	छ्स	छ्ह
ज्	ज्प	ज्फ	ज्ब	ज्भ	ज्म	ज्य	ज्र	ज्ल	ज्व	ज्श	ज्ष	ज्स	ज्ह
झ्	झ्प	झ्फ	झ्ब	झ्भ	झ्म	झ्य	झ्र	झ्ल	झ्व	झ्श	झ्ष	झ्स	झ्ह
ञ्	ञ्प	ञ्फ	ञ्ब	ञ्भ	ञ्म	ञ्य	ञ्र	ञ्ल	ञ्व	ञ्श	ञ्ष	ञ्स	ञ्ह
ट्	ट्प	ट्फ	ट्ब	ट्भ	ट्म	ट्य	ट्र	ट्ल	ट्व	ट्श	ट्ष	ट्स	ट्ह
ठ्	ठ्प	ठ्फ	ठ्ब	ठ्भ	ठ्म	ठ्य	ठ्र	ठ्ल	ठ्व	ठ्श	ठ्ष	ठ्स	ठ्ह
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
ण्	ण्प	ण्फ	ण्ब	ण्भ	ण्म	ण्य	ण्र	ण्ल	ण्व	ण्श	ण्ष	ण्स	ण्ह
त्	त्प	त्फ	त्ब	त्भ	त्म	त्य	त्र	त्ल	त्व	त्श	त्ष	त्स	त्ह
थ्	थ्प	थ्फ	थ्ब	थ्भ	थ्म	थ्य	थ्र	थ्ल	थ्व	थ्श	थ्ष	थ्स	थ्ह
द्	द्प	द्फ	द्ब	द्भ	द्म	द्य	द्र	द्ल	द्व	द्श	द्ष	द्स	द्ह
ध्	ध्प	ध्फ	ध्ब	ध्भ	ध्म	ध्य	ध्र	ध्ल	ध्व	ध्श	ध्ष	ध्स	ध्ह
न्	न्प	न्फ	न्ब	न्भ	न्म	न्य	न्र	न्ल	न्व	न्श	न्ष	न्स	न्ह
प्	प्प	प्फ	प्ब	प्भ	प्म	प्य	प्र	प्ल	प्व	प्श	प्ष	प्स	प्ह
फ्	फ्प	फ्फ	फ्ब	फ्भ	फ्म	फ्य	फ्र	फ्ल	फ्व	फ्श	फ्ष	फ्स	फ्ह

ब्	ब्प	ब्फ	ब्ब	ब्भ	ब्म	ब्य	ब्र	ब्ल	ब्व	ब्श	ब्ष	ब्स	ब्ह
भ्	भ्प	भ्फ	भ्ब	भ्भ	भ्म	भ्य	भ्र	भ्ल	भ्व	भ्श	भ्ष	भ्स	भ्ह
म्	म्प	म्फ	म्ब	म्भ	म्म	म्य	म्र	म्ल	म्व	म्श	म्ष	म्स	म्ह
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	ल्श	ल्ष	ल्स	ल्ह
व्	व्प	व्फ	व्ब	व्भ	व्म	व्य	व्र	व्ल	व्व	व्श	व्ष	व्स	व्ह
श्	श्प	श्फ	श्ब	श्भ	श्म	श्य	श्र[35]	श्ल	श्व	श्श	श्ष	श्स	श्ह
ष्	ष्प	ष्फ	ष्ब	ष्भ	ष्म	ष्य	ष्र	ष्ल	ष्व	ष्श	ष्ष	ष्स	ष्ह
स्	स्प	स्फ	स्ब	स्भ	स्म	स्य	स्र	स्ल	स्व	स्श	स्ष	स्स	स्ह
ह्	ह्प	ह्फ	ह्ब	ह्भ	ह्म	ह्य	ह्र	ह्ल	ह्व	ह्श	ह्ष	ह्स	ह्ह
क़्	क़्प	क़्फ	क़्ब	क़्भ	क़्म	क़्य	क़्र	क़्ल	क़्व	क़्श	क़्ष	क़्स	क़्ह
ख़्	ख़्प	ख़्फ	ख़्ब	ख़्भ	ख़्म	ख़्य	ख़्र	ख़्ल	ख़्व	ख़्श	ख़्ष	ख़्स	ख़्ह
ग़्	ग़्प	ग़्फ	ग़्ब	ग़्भ	ग़्म	ग़्य	ग़्र	ग़्ल	ग़्व	ग़्श	ग़्ष	ग़्स	ग़्ह
ज़्	ज़्प	ज़्फ	ज़्ब	ज़्भ	ज़्म	ज़्य	ज़्र	ज़्ल	ज़्व	ज़्श	ज़्ष	ज़्स	ज़्ह
फ़्	फ़्प	फ़्फ	फ़्ब	फ़्भ	फ़्म	फ़्य	फ़्र	फ़्ल	फ़्व	फ़्श	फ़्ष	फ़्स	फ़्ह
ड़्	ड़्प	ड़्फ	ड़्ब	ड़्भ	ड़्म	ड़्य	ड़्र	ड़्ल	ड़्व	ड़्श	ड़्ष	ड़्स	ड़्ह
ढ़्	ढ़्प	ढ़्फ	ढ़्ब	ढ़्भ	ढ़्म	ढ़्य	ढ़्र	ढ़्ल	ढ़्व	ढ़्श	ढ़्ष	ढ़्स	ढ़्ह

[35] An important combination very often mistyped is श+ृ = शृ which is often mis-written as श+र+ृ= श्रृ Thus शृंगार is wrongly written as श्रृंगार.

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 e.g. The combination of consonants "क" and " क़" is ligature "क्क़".

CHC Set 4						
	क़	ख़	ग़	ज़	ड़	ढ़
क्	क्क़	क्ख़	क्ग़	क्ज़	क्ड़	क्ढ़
ख्	ख्क़	ख्ख़	ख्ग़	ख्ज़	ख्ड़	ख्ढ़
ग्	ग्क़	ग्ख़	ग्ग़	ग्ज़	ग्ड़	ग्ढ़
घ्	घ्क़	घ्ख़	घ्ग़	घ्ज़	घ्ड़	घ्ढ़
ङ्	ङ्क़	ङ्ख़	ङ्ग़	ङ्ज़	ङ्ड़	ङ्ढ़
च्	च्क़	च्ख़	च्ग़	च्ज़	च्ड़	च्ढ़
छ्	छ्क़	छ्ख़	छ्ग़	छ्ज़	छ्ड़	छ्ढ़
ज्	ज्क़	ज्ख़	ज्ग़	ज्ज़	ज्ड़	ज्ढ़
झ्	झ्क़	झ्ख़	झ्ग़	झ्ज़	झ्ड़	झ्ढ़
ञ्	ञ्क़	ञ्ख़	ञ्ग़	ञ्ज़	ञ्ड़	ञ्ढ़
ट्	ट्क़	ट्ख़	ट्ग़	ट्ज़	ट्ड़	ट्ढ़
ठ्	ठ्क़	ठ्ख़	ठ्ग़	ठ्ज़	ठ्ड़	ठ्ढ़
ड्	ड्क़	ड्ख़	ड्ग़	ड्ज़	ड्ड़	ड्ढ़
ढ्	ढ्क़	ढ्ख़	ढ्ग़	ढ्ज़	ढ्ड़	ढ्ढ़
ण्	ण्क़	ण्ख़	ण्ग़	ण्ज़	ण्ड़	ण्ढ़
त्	त्क़	त्ख़	त्ग़	त्ज़	त्ड़	त्ढ़
थ्	थ्क़	थ्ख़	थ्ग़	थ्ज़	थ्ड़	थ्ढ़
द्	द्क़	द्ख़	द्ग़	द्ज़	द्ड़	द्ढ़
ध्	ध्क़	ध्ख़	ध्ग़	ध्ज़	ध्ड़	ध्ढ़
न्	न्क़	न्ख़	न्ग़	न्ज़	न्ड़	न्ढ़
प्	प्क़	प्ख़	प्ग़	प्ज़	प्ड़	प्ढ़
फ्	फ्क़	फ्ख़	फ्ग़	फ्ज़	फ्ड़	फ्ढ़

ब्	ब़क	ब़ख	ब़ा	ब़ज	ब़ड़	ब़ढ़
भ्	भ़क	भ़ख	भ़ा	भ़ज	भ़ड़	भ़ढ़
म्	म़क	म़ख	म़ा	म़ज	म़ड़	म़ढ़
य्	य़क	य़ख	य़ा	य़ज	य़ड़	य़ढ़
र्	ऱक	ऱख	ऱा	ऱज	ऱड़	ऱढ़
ल्	ल़क	ल़ख	ल़ा	ल़ज	ल़ड़	ल़ढ़
व्	व़क	व़ख	व़ा	व़ज	व़ड़	व़ढ़
श्	श़क	श़ख	श़ा	श़ज	श़ड़	श़ढ़
ष्	ष़क	ष़ख	ष़ा	ष़ज	ष़ड़	ष़ढ़
स्	स़क	स़ख	स़ा	स़ज	स़ड़	स़ढ़
ह्	ह़क	ह़ख	ह़ा	ह़ज	ह़ड़	ह़ढ़
क़्	क़क	क़ख	क़ा	क़ज	क़ड़	क़ढ़
ख़्	ख़क	ख़ख	ख़ा	ख़ज	ख़ड़	ख़ढ़
ग़्	ग़क	ग़ख	ग़ा	ग़ज	ग़ड़	ग़ढ़
ज़्	ज़क	ज़ख	ज़ा	ज़ज	ज़ड़	ज़ढ़
फ़्	फ़क	फ़ख	फ़ा	फ़ज	फ़ड़	फ़ढ़
ड़्	ड़क	ड़ख	ड़ा	ड़ज	ड़ड़	ड़ढ़
ढ़्	ढ़क	ढ़ख	ढ़ा	ढ़ज	ढ़ड़	ढ़ढ़

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 mātrās have not been taken into account.[36]

[36] Quite a few examples are culled from Bahri and Nalanda Dictionaries. Others are taken form the CIIL corpus.

CHCHC	Example
क्ष्प	दृक्ष्पेप
क्ट्र	इलेक्ट्रोन
क्ट्स	स्पैक्ट्स
क्त्य	युक्त्याभास
क्त्र	पंचवक्त्र
क्त्व	पृथक्त्व
क्प्र	वाक्प्रतोद
क्ल्य	शौक्ल्य
ष्ण	तीष्ण
ष्म	भाग्यलष्मी
ष्य	अनपेष्य
ष्व	इष्वाकु
क्स्ड	फिक्स्ड
क्स्त	वाक्स्तंभ
क्स्व	वाक्स्वातंत्र्य
ख्व्व	ख्व्वाजा
ग्र	वणिग्ग्राम
ग्ज्य	प्राग्ज्योतिष
ग्द्ध	प्राग्द्धार
ग्न्य	अग्न्याधान
ग्भ्र	दिग्भ्रम
ग्व्य	वाग्व्यवहार
च्छ्व	उच्छ्वास
ज्ज्य	उज्ज्वल

त्क्ल	उत्क्लेश
त्ष्ण	तात्ष्णिक
त्थ	सुप्तोत्थित
त्त्व	तत्त्व
त्प्र	उत्प्रेरक
त्प्ल	उत्प्लव
त्म्य	स्थानमहात्म्य
त्र्य	वैचित्र्य
त्स्थ	तत्स्थानीय
त्स्न	जोत्स्ना
त्स्प	ईषत्स्पृष्ट
त्स्प	हृत्स्पंदों
त्स्फ	उत्स्फूर्त
त्स्य	मत्स्य
त्स्व	तत्स्वरूप
द्ग्र	उद्ग्रीव
द्ज्ञ	सद्ज्ञान
द्द्ध	एतद्द्धारा
द्द्ष	नामोद्द्ष्ट
द्भ्र	उद्भ्रम
द्र्य	दारिद्र्य
द्व्य	सद्व्यवहारों
न्क्व	इन्क्वायरी
न्ट्र	इन्फ़ैन्ट्रियाँ
न्ट्स	स्टूडन्ट्स

न्त्य	अचिन्त्य
न्त्र	उप-मन्त्री
न्द्य	अनिन्द्य
न्द्र	अंतरिक्षकेन्द्र
न्द्र	उपेन्द्रनाथ
न्ध्य	फलवन्ध्य
न्न्य	वैभिन्न्य
न्फ्र	कान्फ्रेंस
प्त्य	प्राप्त्याशा
फ्ट्स	एयरक्राफ्ट्स
ब्ज्य	कौब्ज्य
म्म्र	सम्म्राज
र्क्य	सर्क्यूलरों
र्क्र	अंतर्क्रिया
र्क्स	मार्क्सवाद
र्ख्य	मौर्ख्य
र्ग्य	वर्ग्य
र्ग्र	अंतर्ग्रस्त
र्घ्य	अर्घ्य
र्च्छ	मूर्च्छित
र्च्य	अर्च्य
र्ज्ञ	अंतर्ज्ञान
र्ज्य	वर्ज्य
र्ट्र	पोर्ट्रेट
र्ट्स	स्पोर्ट्समैन

र्ण्य	वर्ण्य
र्त्य	अतिमर्त्य
र्त्र	सार्त्र
र्त्स	वर्त्स
र्थ्य	सामर्थ्यों
र्द्ध	आर्द्ध
र्द्य	सौहार्द्य
र्द्र	आर्द्र
र्द्ध	अंतर्द्वंद्व
र्ध्द	पूर्वार्ध्द
र्ध्य	अंतर्ध्यान
र्ध्व	उर्ध्व
र्प्य	समर्प्य
र्प्र	अंतर्प्रादेशिक
र्फ्य	कर्फ्यू
र्भ्र	निर्भ्रांत
र्म्य	हर्म्य
र्ल्ड	फ्रीवर्ल्ड
र्ल्स	चार्ल्स
र्व्य	बहिर्व्यसनी
र्व्र	निर्व्रण
र्श्व	पार्श्व
र्ष्ण	कार्ष्ण
र्ष्य	ईर्ष्या
र्स्ट	फ़र्स्ट

ल्क्य	याज्ञवल्क्य
ल्ट्र	पोल्ट्रीफार्म
ल्ड्र	चिल्ड्रन
ल्प्य	प्रकल्प्य
ल्भ्य	प्रागल्भ्य
श्व्र	पारिपाश्व्रक
ष्क्र	निष्क्रिय
ष्क्व	निष्क्वाथ
ष्ट्य	उक्तिवैशिष्ट्य
ष्ट्र	अंतरराष्ट्रीय
ष्ठ्य	ओष्ठ्य
ष्ण्य	वैतृष्ण्य

ष्म्र	निष्म्रयोजन
स्क्र	स्क्रिप्ट
स्क्व	मस्क्वा
स्क्व	स्क्वाड्रन
स्ट्र	इंडस्ट्रियल
स्त्य	अगस्त्य
स्त्र	अधोवस्त्र
स्त्व	अंतस्त्वचा
स्थ्य	स्वास्थ्य
स्प्र	स्प्रिंग
स्प्ल	एस्प्लेनेड

6.3.3.3.CHCHCHC (Combination of four Consonants)

These combinations are rare and found only in the following cases[37]:

द्व्र सद्व्रत
स्त्र्य स्त्र्याजीव

6.3.3.4 CHCHCHCHC (Combination of five Consonants)

Apart from the Sanskrit कात्स्न्र्य no other case of a five consonant cluster seems to exist.

[37] The examples are culled from Bahri and Nalanda Dictionaries.

6.3.4. Valid Combinations And Invalid Combinations

As mentioned in the *Introduction* as well as Section 1. *End Users of the "Devanagari Script behaviour for Hindi"* (cf. supra), the document targets basically the font designer desirous of knowing the exact shapes both of individual characters as well as combinations such as
Consonant-Mātrā,
Consonant-Mātrā-Nasal,
Consonant+Halant+Consonant,
Consonant+Halant+Consonant+Halant+Consonant.
It has been therefore decided to display all and every combination both valid and invalid in the document.

6.3.5 The Collation Order of Hindi.

Collation is one of the most important features of this document. 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 document.

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	ऍ[38]	ऐ	ओ	ऑ
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	ऍ
ऐ	ओ	ऑ	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	ा	ि	ी	ु	ू	ृ	े	ॅ	ै	ो	ॉ
ौ	्	़									

[38] ऍ and ऑ 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	ऍ
ऐ	ओ	ऑ	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	ा	ि	ी	ु	ू	ृ	े	ॅ	ै	ो	ॉ
ौ	्	़									

Following is the example of sort order for consonant " क "

कँ	कं	कः	का	कि	की	कु	कू	कृ	के	कॅ	कै
को	कॉ	कौ	क्	क़							

7. REFERENCES

1. http://www.unicode.org
2. INDIAN SCRIPT CODE FOR INFORMATION INTERCHANGE - ISCII IS 13194. 1991. NEW DELHI
3. केंद्रीय हिंदी निदेशालय: देवनागरी लिपि तथा हिंदी वर्तनी का मानकीकरण नई दिल्ली. 2006
4. Omkar N Koul *Modern Hindi Grammar*. Hyattsville: Dunwoody Press, 2008, pp. x+318
5. Yamuna Kachru. *Hindi*. London Oriental and African Language Library, 12 2006. xxii, pp. 309
6. Kalyan Kale, Anjali Soman. *Learning Marathi* 1st ed. Shri Vishakha Prakashan, 1986.

8. ANNEXURES

Annexure – I : Names of experts who have redrafted the document based on inputs of the committee

Prepared by :

- Dr. Raiomond Doctor, Consultant, C-DAC, GIST, Pune

Shapes in this documents are based on CHD book

Inputs by :

- Mr. Prabhakar Pandey, Hindi Officer, C-DAC, Pune
- Mr. Tukaram Patil, Department of Hindi, University of Pune
- Dr. Sudhir Mishra, Hindi Linguist, C-DAC, Pune
- Hindi Language Expert nominated by Director, CIIL

The first meeting of the "Expert Committee on Linguistic Resources and Language Technology Standards" for Validation of Hindi Script Grammar was held at DeitY on 12th Oct, 2012, under the chairmanship of Ms. Swaran Lata, HoD (HCC).

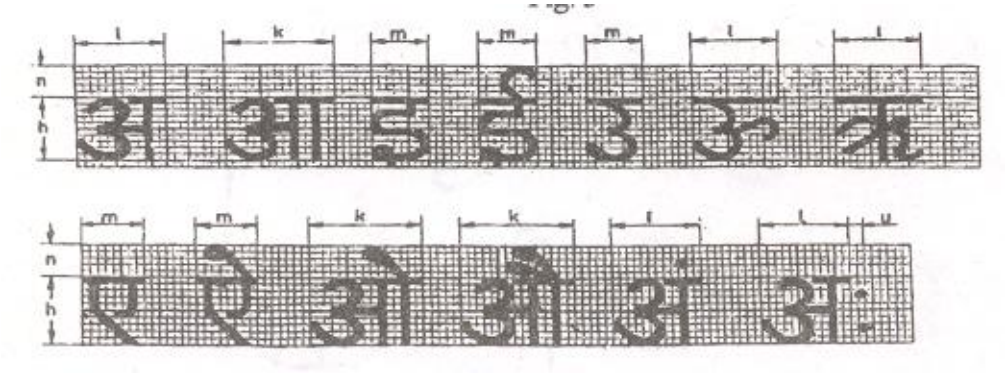
The list of participants :

S. No.	Name	Organization	Mobile	Email
1.	Ms. Swaran Lata	DeitY	9868851170	slata@mit.gov.in
2.	Dr. Mukul Kumar Sinha	Expert Software consultants	9810874961	mukuls@yahoo.co.in
3.	Prof. Panchanan Mohanty	University of Hyderabad	9440034166	panchananmotianty@gmail.com
4.	Prof. B. Mallikarjun	Central University or Karnataka	9449178768	mallikarjun56@gmail.com
5.	Prof. Ghanashyam Nepal	University of North Bengal	9434137452	nepalgs@gmail.com
6.	Sh. Atiur Rahman Khan	CDAC, Pune	9049717192	atiurk@gmail.com
7.	Sh. Prabhakar Pandey	CDAC, Pune	9702127182	prabhakarp@cdac.in
8.	Prof. Hidam Dolen Singh	NERLC	9957115106	dikebgu@gmail.com
9.	Dr. Bhaskarjyoti Sarma	Dibrugarh University	9854514507	bhasdu08@gmail.com
10.	Sh . J.Kumar	NIC Delhi	9868202772	kendolji@nic.in
11.	Prof. Bisweswar Basumatary	Bodo Sahitya Sabha	08011395142	bssjb@rediffmail.com

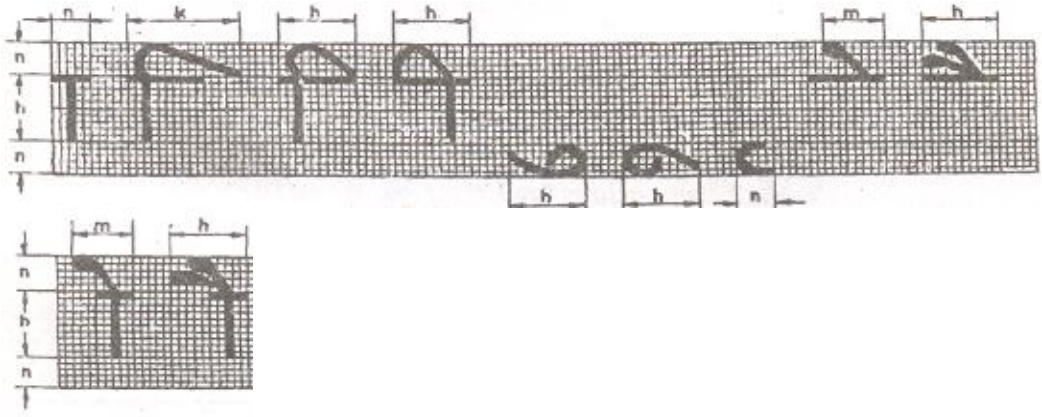
12.	Sh. Manoj Jain	DeitY	9868869196	mjain@mit.gov.in
13.	Sh. Deepak Pandey	CHD	9810722080	dkp410@gmail.com
14.	Ms. Nutan Pandey	CHD	9968284910	ilove.india2008@indiatimes.com
15.	Ms. Richa	LDC-IL, CIIL Mysore	9590426049	rsrishti@gmail.com
16.	Prof. Girish Nath Jha	JNU	9899107988	girishjha@gmail.com
17.	Dr. Sudhir K. Mishra	CDAC, Pune	9763401733	smishra@cdac.in
18.	Sh Pravin Satpute	Redhat	9881038402	psatpute@redhat.com
19.	Prof K. K.Goswami	Retd	9971553740	
20.	Dr. Saurobh Gupta	NIC	9013132222	Saurabh.gupta@nic.in
21.	Sh R.K. Pathak	CTB, DOL, MHA	9013289908	Rakesh.pathak@nic.in
22.	Sh. Kewal Krishan	NIC, D/O official language	9810031413	kewal.krishan@nic.in
23.	Sh. Harinder Kumar	DOL, MHA	8527322229	dol-tech@nic.in
24.	Sh. Vijay Kumar	DeitY, Delhi	9899428992	vkumar@mit.gov.in
25.	Sh. M D Kulkarni	CDAC, Pune	9822053779	mdk@cdac.in
26.	Sh. Jasjit Singh	CDAC	9910163863	jasjit@cdac.in
27.	Sh. Naitik Tyagi	W3C	9999984648	tyagi@w3.org
28.	Ms. Tanushree Ojha	W3C	9555842015	tanu.ojha@gmail.com
29.	Prof. N Pramodni Devi	Manipur University	09862124036	npdini@yahoo.co.in
30.	Prof. Shikhar Kumar Sharma	Gauhati University	9954712669	sks001@gmail.com
31.	Prof. V R Jagannathan	Retd	9871824742	vrj.nathan@gmail.com
32.	Prof. Pushpak Bhattatacharya	IITB, Mumbai	9819708718	pb@cse.iitb.ac.in
33.	Ms. Kiran singh	CDAC, Delhi	9810308396	kirans@cdac.in
34.	Sh. Abhijit Bhattacharjee	Luna Ergonomics	8860030141	abhijit@paninikeyped.com

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

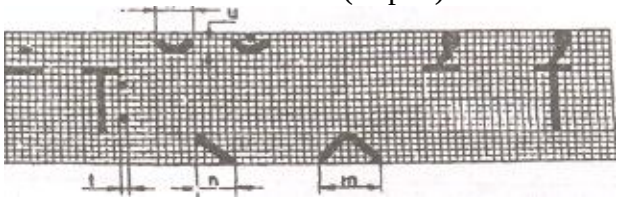
Hindi Vowels



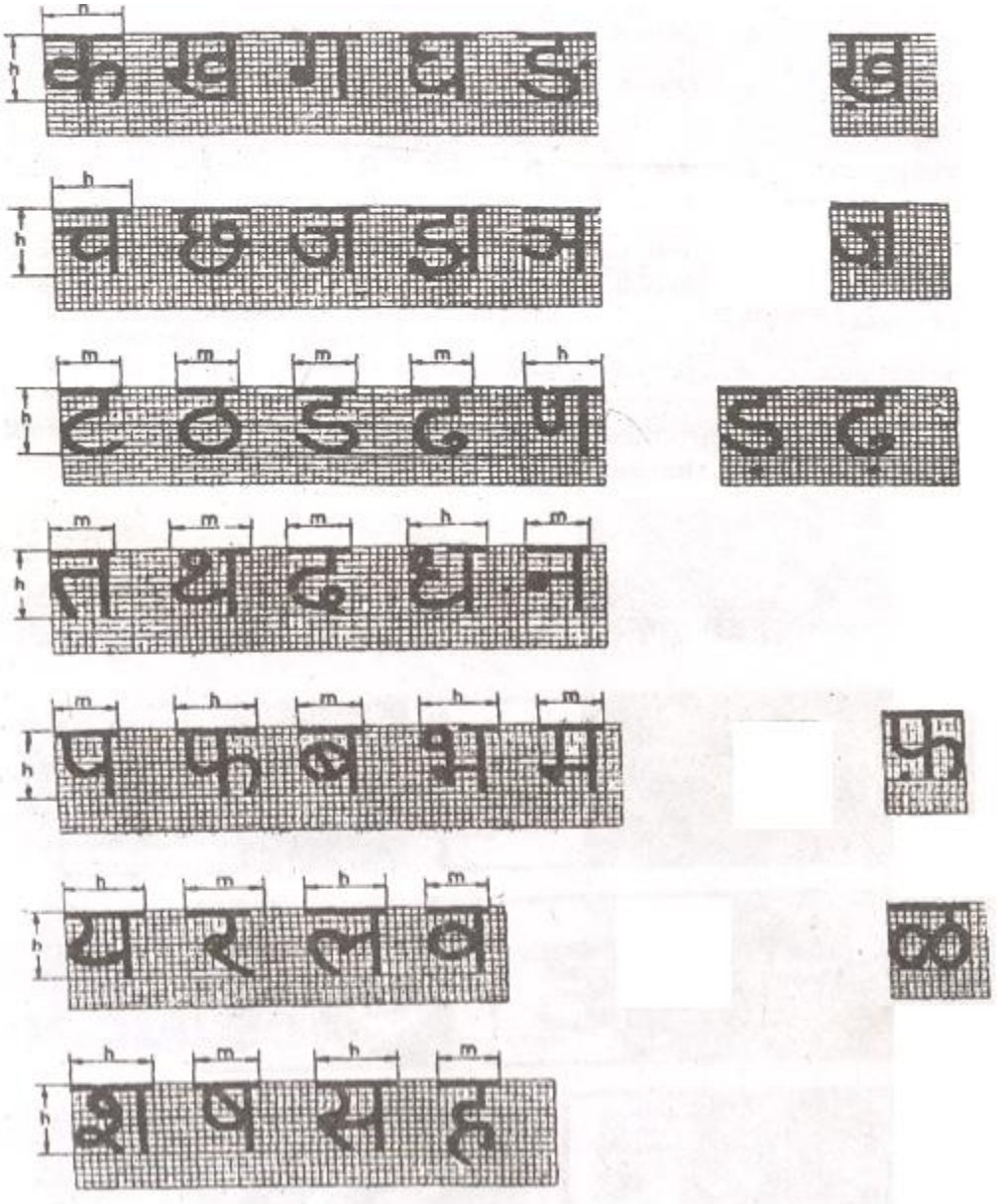
Hindi Vowel Modifiers: Mātrās



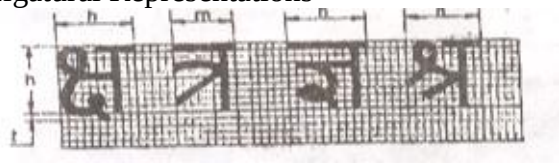
Hindi Diacritics and Rafar(Repha) Rakar



Hindi Consonants



Ligatural Representations



Annexure 3: A Note on Zero Width Joiner and Zero Width Non-Joiner

This note is pertinent to the use of Zero Width Joiner (ZWJ) and Zero Width Non Joiner (ZWNJ) as used in Hindi. It needs to be noted that Nepali, Konkani and Marathi use these two signs in a different manner.

ZWJ (U+0200D) and ZWNJ (U+0200C) are code points that have been provided by the Unicode standard to instruct the rendering of a string where the script has the option between joining and non-joining characters. Without the use of these control codes, the string may be rendered in an alternate form from what is intended.

In the case of Hindi, ZWJ/ZWNJ do not play an important role insofar as functionality is concerned. Unicode 2.0 prescribes the use of RA+VIRAMA+ZWJ to represent the eyelash-ra. This is captured in what was then rule R5 of Section 9 (which is now rule R5a). Unicode 3.0/4.0 reflected the ISCII choice, in what is now rule R5: "*In conformance with the ISCII standard, the half-consonant form rrah is represented as eyelash-ra. This form of ra is commonly used in writing Marathi…*" (Unicode 3.0)

The word द‍र्‍या *daryā* can be written with the Unicode values

U+0926 U+0930 U+094D U+200D U+092F U+093E (द‍र्‍या)

as well as

U+0926 U+0931 U+094D U+092F U+093E (द‍र्‍या)

Insofar as Hindi is concerned ZWJ/ZWNJ are used to render alternate rendering of ligatures.

If a consonant+halant is followed by the ZWJ, the half-form of the consonant is formed. Example:

शक्ति[39] U+0936 U+0915 U+094D U+0924 U+093F

श‍क्ति U+0936 U+0915 U+094D U+200D U+0924 U+093F

अक्षय U+0905 U+0915 U+094D U+0937 U+092F

अक्‍षय U+0905 U+0915 U+094D U+200D U+0937 U+092F

This use of ZWJ serves a pedagogical purpose in that it allows the learner to study and master the half shapes of characters.

[39] As in the case of Marathi

The use of ZWNJ in Hindi is restricted to representing a dead consonant within a string. Thus to show the combination of राज्+कमल as a single word and retain the shape of the consonant followed by the halant; ZWNJ is used:

राज्‌कमल :U+0930 U+093E U+091C U+094D U+200C U+0915 U+092E U+0932

This practice is followed to represent Sanskrit loan words or proper names demanding a "dead" consonant:

उद्‌घोष

A secondary use of ZWJ/ ZWNJ is to reduce a stacked ligature to a linear ligature. Thus:

Without ZWNJ	पक्का in Marathi पक्का in Hindi
With ZWNJ	पक्‌का
With ZWJ	पक्का

As can be seen ZWJ/ZWNJ are mainly used to render alternate forms.

The use of ZWJ/ZWNJ is not permitted in Internationalized Domain Names. Used, as in the case of Hindi, to create alternate renderings, the insertion of these two signs can affect searching as well as NLP.

Annexure 4: Unicode Table of Devanāgarī [40]

[40] The Unicode chart provided is for version 5.1 since the document was prepared at that time. No considerable change in the document can be seen in the updated versions of Unicode, with the possible addition of the Rupee Sign U+20B9. Permission for including the chart is awaited from Unicode and hence the chart is not included in this document.

Annexure 5: Definition of the Indic Akshar

Definition of the Indic Akṣara

V[m]|{C[N]H}C[N](H|[v][m])

Akṣara may best be defined as “Indic syllable”. Indian scripts are directly based on phonetics-the units of orthography exhibit a more or less one to one correspondence with the spoken sounds. The Akṣara represents a vowel, consonant or a conjunct consonant.

Where V (upper case) is any independent vowel, *m* is any vowel modifier (Devanāgari Anusvāra, Visarga, Candrabindu), C is any consonant (with inherent vowel), N is Nukta, H is halant or Virāma and *v* (lower case) is any dependent vowel sign (mātrā). Following Conventions are used in the syntax: **{}** is enclose items which may be repeated one or more times, **[ ]** is enclose items which may not be present and **|** is separates items, out of which only one can be present.
The definition given above in the form of a regular expression may be paraphrased as follows:

- any independent vowel is an akṣara, e.g.
  अ, ई, उ
- any independent vowel followed by anusvāra, candrabindu or visarga is an akṣara, e.g.

  अं, उँ, आः
- zero or more consonant(+Nukta)+virāma sequences followed by a consonant (+Nukta) is an akṣara, e.g.
  र्त, र्त्स, र्त्स्न, र्त्स्न्य, फ़्क़, वफ़्क़
- zero or more consonant(+Nukta)+virāma sequences followed by a consonant (+Nukta) followed by a virāma is an akṣara, e.g.
  र्त्, स्त्,फ़्क़्
- zero or more consonant+(Nukta)+virāma sequences followed by a consonant (+Nukta) followed by a vowel sign is an akṣara, e.g.

र्ता, त्स्न्र्या, फ़ज़ी,लफ़ज़ी

• zero or more consonant+(Nukta)+virāma sequences followed by a consonant (+Nukta) followed by an anusvāra or candrabindu or visarga is an akṣara, e.g.

स्तं, स्त्रँ, स्तः, फ़ज़ँ

• zero or more consonant+(Nukta)+virāma sequences followed by a consonant (+Nukta) followed by a vowel sign and an anusvāra or candrabindu or visarga is an akṣara, e.g.

त्स्न्र्याः, त्स्न्युं, त्स्न्युँ, फ़ज़ें

• nothing else is an akṣara.