

“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	3
1. OBJECTIVES OF "Devanagari Script behaviour for Hindi"	6
2. END USERS FOR "Devanagari Script behaviour for Hindi"	8
3. SCOPE	9
4. TERMINOLOGY	10
5. PHILOSOPHY AND UNDERLYING PRINCIPLES	14
6. "Devanagari Script behaviour for Hindi" : STRUCTURE	15
6.1. PERIPHERAL ELEMENTS OF THE "Devanagari Script behaviour for Hindi"	16
6.2. CONFORMITY TO THE SYLLABLE STRUCTURE	17
6.3 "DEVANAGARI SCRIPT BEHAVIOUR FOR HINDI" PROPER	21
6.3.1. The Character Set of Hindi.	21
6.3.2. Consonant Mātrā Combinations.	30
6.3.3. The Ligature Set of Hindi.	46
6.3.4. Valid Combinations And Invalid Combinations.....	59
6.3.5 The Collation Order of Hindi.	60
7. REFERENCES	62
8. ANNEXURES	63
Annexure – I : Names of experts who have redrafted the document based on inputs of the committee.....	63
Annexure 2: Shapes of Hindi characters as per Central Hindi Directorate	65
Annexure 3: A Note on Zero Width Joiner and Zero Width Non-Joiner	67
Annexure 4: Unicode Table of Devanāgarī	69
Annexure 5: Definition of the Indic Akshar	70

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 ं and ः (for generating out the eyelash ra). These are not present in Hindi or Dogri. The Hindi ं (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 /ଲ/ /Marathi /ଳ/ .

The same is the case with टि which is represented in Hindi as टि but in Marathi as

ଟି. Numbers also display differences

The Hindi number set is as under:

० १ २ ३ ४ ५ ६ ७ ८ ९

The Marathi and Konkani number set is as under:

० १ २ ३ ४ ५ ६ ७ ८ ९

- 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 linear if the language is Hindi and as stacked if the language is Nepali or Marathi¹

Hindi: शक्ति

Nepali/Marathi: शक्ति

- d. The collation order within the language.

The collation order² varies from language to language although they all share the same script.

In the case of Hindi

श ज त्र are sorted along with the first consonant of each ligature. Thus श is sorted along with क, ज with ज and त्र with त

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

In Nepali श ज 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

¹, Although the Central Hindi Directorate recommends व्त्र as the normative form, in Hindi both versions seem to be used cf. books published by Gorakhpur Press and other reputed publishers (शक्ति & as शक्ति). It would appear that both forms are acceptable in Hindi writing style.

² Cf. 6.3.5 of the document

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. 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 झ, झ़ are placed at the end of the consonant inventory, i.e. after ह in the sort order. In Hindi झ is sorted along with क and झ 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. TERMINOLOGY³

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.⁴ 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 variant⁵. 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 - ḷ [U+ 0933] which Hindi does not use. Flaps used in Hindi ଙ [U+095C] ଙ [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 obstruction⁶. A *consonant, especially a stop consonant followed by a puff of breath that is clearly audible before the next sound begins.*⁷

Example ḷ 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.

³ 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.

⁴ Wikipedia definition

⁵ <http://dictionary.reference.com/browse/allograph>

⁶ Cf. P 1. 2.1.3. of BIS Document 1991.

⁷ <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.⁸

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 /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*: य̄ ह̄. It is used in Konkani, Nepali and Marathi. In Marathi not all combinations of this type generate an *eyelash ra* e.g. द्या॒रा /darya/ “ocean” vs. द्या॒या /darya/ “valleys”. Unicode prescribes a combination of र्॒+य U+0931 U+094D U+092F र्॒+ह̄ U+0931 U+094D U+0939 for generating the *eyelash ra*. Earlier the *eyelash ra* was generated by a combination of *ra+halant+ZWJ र्* i.e. U+0930 U+094D U+200D.^{9 10}

⁸ Unicode ver. 6.0 Chapter 9.0 pp. 6-7.

⁹ <http://unicode.org/~emuller/iwg/p8/utcdoc.html>

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: जँच

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

Pure consonant¹¹: 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 = क̄॑ 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 = घ̄॑. 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. ॲ॑ In the case of consonants which have no stem such as the retroflexes in Devanāgarī, the rakar is placed below the consonant to which it relates e.g. ॲ॒.

¹⁰ Cf. p64 of this document

¹¹ The term used is as per BIS DOCUMENT **IS 13194: 1991**. However, it could also be termed as “short obstrusive sound” as per suggestion of experts.

Repha: (see **Rafar**).

Roman script¹²: 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*¹³. 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*¹⁴.

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).

¹² The term used is as per BIS DOCUMENT **IS 13194: 1991**.**Point No 2.1.20 p. 1**

¹³ Cf. P. 1 2.1.6 BIS Document IS 13194. 1991

¹⁴ <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 allographs 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 occur¹⁵.
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.

¹⁵ 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 Anuswara¹⁶ and Candrabindu are also provided.

¹⁶ Although Unicode uses the term Anuswar and some texts use the term Anunasika, the correct term to be used is Bindu. Anuswar is the linguistic term used for characterizing NASAL VOWELS, and anunasika is the term used for characterizing NASAL CONSONANTS. The diacritics should simply be called 'bindu' and chandrabindu'. (Dr Pramod Pandey)

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 क्॒ष = ख्।

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 STRUCTURE¹⁷

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
ङ	घ	ঁ	চ	ঁ
জ	ঁ	ঝ	অ	ট
ঠ	ঁ	ঢ	ঁ	ঁ
ণ	ত	থ	দ	ধ
ন	প	ফ	ঁ	ব
ঁ	ম	য	ৰ	ল
ও	শ	ষ	স	হ

(V) Vowels^{18 19}

অ	আ	ই	ই	উ	ऊ	ু	ু	ে	ে	ো	ো	ৌ
---	---	---	---	---	---	---	---	---	---	---	---	---

(M) Mātrās

¹⁷ Annexure 5 provides a formalism for the Indic Akshar as represented on the W3CIndia site

¹⁸ অঁ অঃ are taught in primary schools as part of the “barakhadi”. However they have been excluded in the list since these are not individual characters per se but are combinations of the vowel অ and ঁ or ঃ respectively. It is for this reason that Unicode does not provide separate code points for অঁ অঃ.

¹⁹ এঁ ওঁ are accepted for transliterating Loan words borrowed from English such as আঁক. In the case of the two short vowels এঁ, ওঁ used in Dravidian, as per recommendation of Central Hindi Directorate 3.15.6.1.2 p. 35, words having a shrot letter/short vowel sign and admitted in Hindi are automatically replaced by the corresponding long letter/vowel sign. Thus the short “o” of Oddanchatram ଓଡ଼ନ୍ଚାତ୍ରମ୍ is rendered as long o in Hindi: ଓଡ଼ିଶାନ୍ତତ୍ତ୍ୱାରମ୍

া	ি	ঁ	ঽ	ঽ	ঽ	ঁ	ঁ	ঁ	ঁ	ঁ	ঁ
---	---	---	---	---	---	---	---	---	---	---	---

(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 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.
ঁ - Avagraha ²⁰	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)²¹ 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.

²⁰ Avagraha is rarely used in Hindi and is used for representing text in Sanskrit where the character is needed..

²¹ 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 ڭ, ڭ, ڭ, ڭ, ڭ

- 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)

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: इ /i/, 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 : कः /ka/
- 2.2. A consonant²² (with or without Nukta) followed by a mātrā, i.e. the inherent vowel being substituted by another vowel: की /ki:/, की /qi:/
- 2.3. A consonant (with or without Nukta) followed by a modifier: कं /kə/, हः /haH²³/
- 2.4. A consonant (with or without Nukta) followed by a mātrā and a modifier: कुं /kū/, दुं /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 /tkə/, त्कः /tkaH/ त्कुं /tkū/, त्दु /tdu/.

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

This means that theoretically the following forms can be postulated:

²² For purposes of Simplification, C here will automatically be treated as being also consonant+nukta: C+N

²³ This character represents phonetically the weak implicit vowel, termed as schwa and often shown as /a/ also.

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 ²⁴	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.

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.

²⁴ C here will automatically be treated as being also consonant+nukta, C+N to simplify the explanation

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.²⁵
- 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.

²⁵ 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 Vargas²⁶.

	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	ड़			स्त्र		
Palato-alveolar								श		
Palatal	च	छ	ज	झ	ज					य
Retroflex	ट	ठ	ड	ढ	ण			ष		
Flaps			ड़	ढ़						
Dental/Alveolar ²⁷	त	थ	द	ধ	ন	র	ল	স	়জ	
Labio-Dental								ফ		ব
Bi-labial	প	ফ	ব	ভ	ম					

Note: Ligatures क্ষ ত্র জ্ঞ শ্র are not listed in the consonants list, since they are

ligatures²⁸: क্ষ=ক+়+ষ, জ্ঞ=জ+়+়জ, ত্র=ত+়+ৰ, শ্র=়ৱ+়ৰ.

However in standard textbooks as per CBSE syllabus these ligatures are treated as individual consonants and not ligatural forms. The Central Hindi Directorate clearly recommends that these be treated as ligatural forms and not individual consonants.

²⁶ 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..

²⁷ The two points of articulation are placed together for sake of economy.

²⁸ 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								ਹ		
Uvular	ਕੁ		ਗ੍ਰ							
Velar	ਕ	ਖ	ਗ	ਧ	ਡ			ਖ		
Palato-alveolar								ਸ਼		
Palatal	ਚ	ਛ	ਜ	ਝ	ਯ					ਯ
Retro-flex	ਟ	ଠ	ଡ	ଢ	ଣ			ସ		
Flaps			ଡ	ଢ						
Dental/Alveolar	ਤ	ਥ	ਦ	ਧ	ਨ	ର	ଲ	ਸ	ଜ	
Labio-Dental								ଫ		ଵ
Bi-labial	ਪ	ਫ	ਬ	ਭ	ਮ					

6.3.1.2. The Vowel Set

The Vowel set of Hindi is as under:

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

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

অ আ ই ঈ উ ঊ ঋ এ ঐ ঃ ও ঔ

6.3.1.3. The Mātrā Set

The Mātrā²⁹ (Vowel Sign) of Hindi is as under:

Mātrā Names	Mātrās Sign	Where is it used ?	Consonant Shapes formed
DEVANAGARI VOWEL SIGN AA	ॐ	आ	क् + आ = का
DEVANAGARI VOWEL SIGN I	ि	इ	क् + इ = कि
DEVANAGARI VOWEL SIGN II	ी	ई	क् + ई = की
DEVANAGARI VOWEL SIGN U	ঁ	উ	ক্ + উ = কু
DEVANAGARI VOWEL SIGN UU	ঁ	ऊ	ক্ + ঊ = কূ
DEVANAGARI VOWEL SIGN VOCALIC R	ঁ	ঁ	ক্ + ঁ = কু
DEVANAGARI VOWEL SIGN E	ঁ	এ	ক্ + এ = কে
DEVANAGARI VOWEL SIGN CANDRA E	ঁ	ঁ	ক্ + ঁ = কঁ
DEVANAGARI VOWEL SIGN AI	ঁ	ঁ	ক্ + ঁ = কঁ
DEVANAGARI VOWEL SIGN O	ঁ	ও	ক্ + ও = কো
DEVANAGARI VOWEL SIGN CANDRA O	ঁ	ঁ	ক্ + ঁ = কঁ
DEVANAGARI VOWEL SIGN AU	ঁ	ঁ	ক্ + ঁ = কঁ

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

ঁ ি ঀ ঁ ঁ ঁ ঁ ঁ ঁ ঁ ঁ

²⁹ 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
ஃ	To left of character	କି, ରି, ପି

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: कं तं e.g. top rafars will be formed in case of following words.

धर्म, चर्खा

RAKARS

- | | | |
|-----------------|-----|-----|
| 1. Bottom rakar | ट | झ |
| 2. Side rakar | क्र | च्र |
| 3. Inside rakar | ह | |

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

Bottom rakar	झम, राष्ट्र
Side rakar	व्रत, चक्र, प्रसाद, स्रोत
Inside rakar	हस्य, हास

6.3.1.6. Diacritics

These are as under in the case of Hindi:

ঁ - Anuswar রংগ

ঁ - Candrabindu/Anunasika ঝঁট

ঁ - Visarga দুর্খ

ঁ - For extra length with long vowels e.g. /উপদেশেওজনুনাসিক /³⁰

6.3.1.7. Halant

ঁ - Halant সংচত্

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.

০ ১ ২ ৩ ৪ ৫ ৬ ৭ ৮ ৯

Numeral Shapes	Explanation
০	Devanāgarī Digit Zero
১	Devanāgarī Digit One
২	Devanāgarī Digit Two
৩	Devanāgarī Digit Three
৪	Devanāgarī Digit Four
৫	Devanāgarī Digit Five
৬	Devanāgarī Digit Six
৭	Devanāgarī Digit Seven
৮	Devanāgarī Digit Eight
৯	Devanāgarī Digit Nine

³⁰ 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 कछु नहीं चाहिए वह शहंशाह ॥
माटी कहे कुम्हार से, तू क्या रौंदे मोय । इक दिन ऐसा आएगा, मैं रौंदँगी तोय ॥

A list of punctuations is provided below:

Devanagari Name	Name of the marker	Marker Shape
"स्टॉप" ³¹	U+002E FULL STOP	.
प्रश्न चिह्न	U+003F QUESTION MARK	?
अल्पविराम	U+002C COMMA	,
विस्मयसूचक चिह्न	U+0021 EXCLAMATION MARK	!
ऊर्ध्व अल्पविराम	U+0027 APOSTROPHE	,
अर्ध्वविराम	U+003B SEMICOLON	;
उपविराम	U+003A COLON	:
योजक चिह्न	U+002D HYPHEN-MINUS	-
निर्देशक चिह्न	U+002D HYPHEN-MINUS U+002D HYPHEN-MINUS Dash	--
लोप चिह्न	FULL STOP (U+002E) FULL STOP (U+002E) FULL STOP (U+002E) Elimination Sign	...
	U+002F SLASH	/
उद्धरण चिह्न	U+0022 QUOTATION MARK	“ ”
शब्द चिह्न	U+0027 APOSTROPHE SINGLE QUOTATION MARK	‘ ’
	U+0058 LATIN CAPITAL LETTER X used as a cross-out character.	X
	U+002D HYPHEN-MINUS U+002D HYPHEN-MINUS U+0022 QUOTATION MARK U+002D HYPHEN-MINUS U+002D HYPHEN-MINUS as above	-- " --

³¹ <http://hi.wikipedia.org/s/ys0>

कोष्ठक	U+0028 LEFT PARENTHESIS U+0029 RIGHT PARENTHESIS	()
कोष्ठक	U+005B LEFT SQUARE BRACKET U+005D RIGHT SQUARE BRACKET	[]
कोष्ठक	U+007B LEFT CURLY BRACKET U+007D RIGHT CURLY BRACKET	{ }
संक्षेपसूचक चिह्न	U+0970 DEVANAGARI ABBREVIATION SIGN	◦
पूरण विराम	U+0964 DEVANAGARI DANDA Devanāgarī Purna Viram	
दीर्घ विराम	U+0965 DEVANAGARI DOUBLE DANDA Devanāgarī Deergha Viram	
हंसपद / त्रुटिबोधक	Sign of left/Add word	↖

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)

¤: 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	ड़	च	छ	ज	झ़	ञ
Table 2:	ट	ठ	ड	ढ	ण	त	थ	द	ধ	ন
Table 3:	প	ফ	ব	ভ	ম	য	ৰ	ল	ৱ	শ
	ষ	স	হ							
Table 4:	ক্	খ্	গ্	ঝ	ফ্	ঢ়	ছ			

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 ଙ + 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: ଙଠ

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	ঙী	চী	ছী	জী	ঝী	ঞী
ু	কু	খু	গু	ঘু	ঙু	চু	ছু	জু	ঝু	ঞু
ূ	কূ	খূ	গূ	ঘূ	ঙূ	চূ	ছূ	জূ	ঝূ	ঞূ
ৃ	কৃ	খৃ	গৃ	ঘৃ	ঙৃ	চৃ	ছৃ	জৃ	ঝৃ	ঞৃ
ে	কে	খে	গে	ঘে	ঙে	চে	ছে	জে	ঝে	ঞে
ঁ	কঁ	খঁ	গঁ	ঘঁ	ঙঁ	চঁ	ছঁ	জঁ	ঝঁ	ঞঁ
ৈ	কৈ	খৈ	গৈ	ঘৈ	ঙৈ	চৈ	ছৈ	জৈ	ঝৈ	ঞৈ
ো	কো	খো	গো	ঘো	ঙো	চো	ছো	জো	ঝো	ঞো
ঁো	কঁো	খঁো	গঁো	ঘঁো	ঙঁো	চঁো	ছঁো	জঁো	ঝঁো	ঞঁো
ৌ	কৌ	খৌ	গৌ	ঘৌ	ঙৌ	চৌ	ছৌ	জৌ	ঝৌ	ঞৌ

Remark 1- ঙ and ঞ are rarely used only as the first members of clusters and mostly as परस्वर्ण or alternatives of अनुस्वार

Consonant and Mātrā combinations Set 2

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

	ਟ	ਠ	ਡ	ਛ	ਣ	ਤ	ਥ	ਦ	ਧ	ਨ
ਾ	ਟਾ	ਠਾ	ਡਾ	ਛਾ	ਣਾ	ਤਾ	ਥਾ	ਦਾ	ਧਾ	ਨਾ
ਿ	ਟਿ	ਠਿ	ਡਿ	ਛਿ	ਣਿ	ਤਿ	ਥਿ	ਦਿ	ਧਿ	ਨਿ
ੀ	ਟੀ	ਠੀ	ਡੀ	ਛੀ	ਣੀ	ਤੀ	ਥੀ	ਦੀ	ਧੀ	ਨੀ
ੁ	ਟੁ	ਠੁ	ਡੁ	ਛੁ	ਣੁ	ਤੁ	ਥੁ	ਦੁ	ਧੁ	ਨੁ
ੂ	ਟੂ	ਠੂ	ਡੂ	ਛੂ	ਣੂ	ਤੂ	ਥੂ	ਦੂ	ਧੂ	ਨੂ
ੈ	ਟੈ	ਠੈ	ਡੈ	ਛੈ	ਣੈ	ਤੈ	ਥੈ	ਦੈ	ਧੈ	ਨੈ
ੌ	ਟੌ	ਠੌ	ਡੌ	ਛੌ	ਣੌ	ਤੌ	ਥੌ	ਦੌ	ਧੌ	ਨੌ
ਾ	ਟੌ	ਠੌ	ਡੌ	ਛੌ	ਣੌ	ਤੌ	ਥੌ	ਦੌ	ਧੌ	ਨੌ
ੰ	ਟੰ	ਠੰ	ਡੰ	ਛੰ	ਣੰ	ਤੰ	ਥੰ	ਦੰ	ਧੰ	ਨੰ
ੈ	ਟੈ	ਠੈ	ਡੈ	ਛੈ	ਣੈ	ਤੈ	ਥੈ	ਦੈ	ਧੈ	ਨੈ
ੋ	ਟੋ	ਠੋ	ਡੋ	ਛੋ	ਣੋ	ਤੋ	ਥੋ	ਦੋ	ਧੋ	ਨੋ
ੁ	ਟੁ	ਠੁ	ਡੁ	ਛੁ	ਣੁ	ਤੁ	ਥੁ	ਦੁ	ਧੁ	ਨੁ
ੌ	ਟੌ	ਠੌ	ਡੌ	ਛੌ	ਣੌ	ਤੌ	ਥੌ	ਦੌ	ਧੌ	ਨੌ

Consonant and Mātrā combinations Set 3

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

	ਪ	ਫ	ਬ	ਭ	ਮ	ਯ	ਰ	ਲ	ਵ	ਸ਼	਷	ਸ	ਹ
ਾ	ਪਾ	ਫਾ	ਬਾ	ਭਾ	ਮਾ	ਯਾ	ਰਾ	ਲਾ	ਵਾ	ਸ਼ਾ	਷ਾ	ਸਾ	ਹਾ
ਿ	ਪਿ	ਫਿ	ਬਿ	ਭਿ	ਮਿ	ਯਿ	ਰਿ	ਲਿ	ਵਿ	ਸ਼ਿ	਷ਿ	ਸਿ	ਹਿ
ੀ	ਪੀ	ਫੀ	ਬੀ	ਭੀ	ਮੀ	ਯੀ	ਰੀ	ਲੀ	ਵੀ	ਸ਼ੀ	਷ੀ	ਸੀ	ਹੀ
ੁ	ਪੁ	ਫੁ	ਬੁ	ਭੁ	ਮੁ	ਯੁ	ਰੁ	ਲੁ	ਵੁ	ਸ਼ੁ	਷ੁ	ਸੁ	ਹੁ
ੂ	ਪੂ	ਫੂ	ਬੂ	ਭੂ	ਮੂ	ਯੂ	ਰੂ	ਲੂ	ਵੂ	ਸ਼ੂ	਷ੂ	ਸੂ	ਹੂ
ੈ	ਪੈ	ਫੈ	ਬੈ	ਭੈ	ਮੈ	ਯੈ	ਰੈ	ਲੈ	ਵੈ	ਸ਼ੈ	਷ੈ	ਸੈ	ਹੈ
ੋ	ਪੋ	ਫੋ	ਬੋ	ਭੋ	ਮੋ	ਯੋ	ਰੋ	ਲੋ	ਵੋ	ਸ਼ੋ	਷ੋ	ਸੋ	ਹੋ
ੌ	ਪੌ	ਫੌ	ਬੌ	ਭੌ	ਮੌ	ਯੌ	ਰੌ	ਲੌ	ਵੌ	ਸ਼ੌ	਷ੌ	ਸੌ	ਹੌ
ੈਂ	ਪੈਂ	ਫੈਂ	ਬੈਂ	ਭੈਂ	ਮੈਂ	ਯੈਂ	ਰੈਂ	ਲੈਂ	ਵੈਂ	ਸ਼ੈਂ	਷ੈਂ	ਸੈਂ	ਹੈਂ
ੋਂ	ਪੋਂ	ਫੋਂ	ਬੋਂ	ਭੋਂ	ਮੋਂ	ਯੋਂ	ਰੋਂ	ਲੋਂ	ਵੋਂ	ਸ਼ੋਂ	਷ੋਂ	ਸੋਂ	ਹੋਂ
ੌਂ	ਪੌਂ	ਫੌਂ	ਬੌਂ	ਭੌਂ	ਮੌਂ	ਯੌਂ	ਰੌਂ	ਲੌਂ	ਵੌਂ	ਸ਼ੌਂ	਷ੌਂ	ਸੌਂ	ਹੌਂ

Consonant and Mātrā combinations Set 4

This set is in continuation of set 3 which shows consonant and mātrā combinations for nukta consonants.³²

	क्र	ख्र	ग्र	ज्ञ	फ्र	ड़्	ठ़्
ा	क्रा	ख्रा	ग्रा	ज्ञा	फ्रा	ड़ा	ठ़ा
ि	क्रि	ख्रि	ग्रि	ज्ञि	फ्रि	ड़ि	ठ़ि
ी	क्री	ख्री	ग्री	ज्ञी	फ्री	ड़ी	ठ़ी
ु	क्रु	ख्रु	ग्रु	ज्ञु	फ्रु	ड़ु	ठ़ु
ू	क्रू	ख्रू	ग्रू	ज्ञू	फ्रू	ड़ू	ठ़ू
ृ ³³	क्रृ	ख्रृ	ग्रृ	ज्ञृ	फ्रृ	ड़ृ	ठ़ृ
े	क्रे	ख्रे	ग्रे	ज्ञे	फ्रे	ड़े	ठ़े
ঁ	ক্রঁ	খ্রঁ	গ্রঁ	জ্ঞঁ	ফ্রঁ	ড়ঁ	ঠ়ঁ
ৈ	ক্রৈ	খ্রৈ	গ্রৈ	জ্ঞৈ	ফ্রৈ	ড়ৈ	ঠ়ৈ
ো	ক্রো	খ্রো	গ্রো	জ্ঞো	ফ্রো	ড়ো	ঠ়ো
ুঁ	ক্রুঁ	খ্রুঁ	গ্রুঁ	জ্ঞুঁ	ফ্রুঁ	ড়ুঁ	ঠ়ুঁ
ৈঁ	ক্রৈঁ	খ্রৈঁ	গ্রৈঁ	জ্ঞৈঁ	ফ্রৈঁ	ড়ৈঁ	ঠ়ৈঁ

³² ক্র, গ্র were not recognized as part of the consonant set by the Central Hindi Directorate, but have been accepted in its latest version..

³³ 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

	ਕ	ਖ	ਗ	ਘ	ਙ	ਚ	ਛ	ਜ	ਝ	ਯ
ਂ	ਕੰ	ਖੰ	ਗੰ	ਘੰ	ਙੰ	ਚੰ	ਛੰ	ਜੰ	ਝੰ	ਯੰ
ਾਂ	ਕਾਂ	ਖਾਂ	ਗਾਂ	ਘਾਂ	ਙਾਂ	ਚਾਂ	ਛਾਂ	ਜਾਂ	ਝਾਂ	ਯਾਂ
ਿਂ	ਕਿਂ	ਖਿਂ	ਗਿਂ	ਘਿਂ	ਙਿਂ	ਚਿਂ	ਛਿਂ	ਜਿਂ	ਝਿਂ	ਯਿਂ
ੀਂ	ਕੀਂ	ਖੀਂ	ਗੀਂ	ਘੀਂ	ਙੀਂ	ਚੀਂ	ਛੀਂ	ਜੀਂ	ਝੀਂ	ਯੀਂ
ੁਂ	ਕੁਂ	ਖੁਂ	ਗੁਂ	ਘੁਂ	ਙੁਂ	ਚੁਂ	ਛੁਂ	ਜੁਂ	ਝੁਂ	ਯੁਂ
ੂਂ	ਕੂਂ	ਖੂਂ	ਗੂਂ	ਘੂਂ	ਙੂਂ	ਚੂਂ	ਛੂਂ	ਜੂਂ	ਝੂਂ	ਯੂਂ
ੈਂ	ਕੈਂ	ਖੈਂ	ਗੈਂ	ਘੈਂ	ਙੈਂ	ਚੈਂ	ਛੈਂ	ਜੈਂ	ਝੈਂ	ਯੈਂ
ੋਂ	ਕੋਂ	ਖੋਂ	ਗੋਂ	ਘੋਂ	ਙੋਂ	ਚੋਂ	ਛੋਂ	ਜੋਂ	ਝੋਂ	ਯੋਂ
ੌਂ	ਕੌਂ	ਖੌਂ	ਗੌਂ	ਘੌਂ	ਙੌਂ	ਚੌਂ	ਛੌਂ	ਜੌਂ	ਝੌਂ	ਯੌਂ
ੈਂਹੈਂ	ਕੈਂਹੈਂ	ਖੈਂਹੈਂ	ਗੈਂਹੈਂ	ਘੈਂਹੈਂ	ਙੈਂਹੈਂ	ਚੈਂਹੈਂ	ਛੈਂਹੈਂ	ਜੈਂਹੈਂ	ਝੈਂਹੈਂ	ਯੈਂਹੈਂ

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
ঁ	ঁ	ঁ	ঁ	ঁ	ঁ	ঁ	ঁ	ঁ	ঁ	ঁ
ঠাঁ	ঠাঁ	ঠাঁ	ঠাঁ	ঠাঁ	ঠাঁ	ঠাঁ	ঠাঁ	ঠাঁ	ঠাঁ	ঠাঁ
ঠিঁ	ঠিঁ	ঠিঁ	ঠিঁ	ঠিঁ	ঠিঁ	ঠিঁ	ঠিঁ	ঠিঁ	ঠিঁ	ঠিঁ
ঠীঁ	ঠীঁ	ঠীঁ	ঠীঁ	ঠীঁ	ঠীঁ	ঠীঁ	ঠীঁ	ঠীঁ	ঠীঁ	ঠীঁ
ঠুঁ	ঠুঁ	ঠুঁ	ঠুঁ	ঠুঁ	ঠুঁ	ঠুঁ	ঠুঁ	ঠুঁ	ঠুঁ	ঠুঁ
ঠুঁ	ঠুঁ	ঠুঁ	ঠুঁ	ঠুঁ	ঠুঁ	ঠুঁ	ঠুঁ	ঠুঁ	ঠুঁ	ঠুঁ
ঠুঁ	ঠুঁ	ঠুঁ	ঠুঁ	ঠুঁ	ঠুঁ	ঠুঁ	ঠুঁ	ঠুঁ	ঠুঁ	ঠুঁ
ঠেঁ	ঠেঁ	ঠেঁ	ঠেঁ	ঠেঁ	ঠেঁ	ঠেঁ	ঠেঁ	ঠেঁ	ঠেঁ	ঠেঁ
ঠঁ ₃₄	ঠঁ									
ঠৈঁ	ঠৈঁ	ঠৈঁ	ঠৈঁ	ঠৈঁ	ঠৈঁ	ঠৈঁ	ঠৈঁ	ঠৈঁ	ঠৈঁ	ঠৈঁ
ঠোঁ	ঠোঁ	ঠোঁ	ঠোঁ	ঠোঁ	ঠোঁ	ঠোঁ	ঠোঁ	ঠোঁ	ঠোঁ	ঠোঁ
ঠোঁ	ঠোঁ	ঠোঁ	ঠোঁ	ঠোঁ	ঠোঁ	ঠোঁ	ঠোঁ	ঠোঁ	ঠোঁ	ঠোঁ
ঠৌঁ	ঠৌঁ	ঠৌঁ	ঠৌঁ	ঠৌঁ	ঠৌঁ	ঠৌঁ	ঠৌঁ	ঠৌঁ	ঠৌঁ	ঠৌঁ

³⁴ ঁ,ঠঁ are used only for the purpose of writing loan words mainly from English. However, since such occurrences are rare, the nasal shapes of ঁ,ঠঁ 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

	ਪ	ਫ	ਬ	ਭ	ਮ	ਯ	ਰ	ਲ	ਵ	ਸ਼	ਬ	ਸ	ਹ
੦	ਪੰ	ਫੰ	ਬੰ	ਭੰ	ਮੰ	ਯੰ	ਰੰ	ਲੰ	ਵੰ	ਸ਼ੰ	ਬੰ	ਸੰ	ਹੰ
ਾਂ	ਪਾਂ	ਫਾਂ	ਬਾਂ	ਭਾਂ	ਮਾਂ	ਯਾਂ	ਰਾਂ	ਲਾਂ	ਵਾਂ	ਸਾਂ	ਬਾਂ	ਸਾਂ	ਹਾਂ
ਿਂ	ਪਿਂ	ਫਿਂ	ਬਿਂ	ਭਿਂ	ਮਿਂ	ਧਿਂ	ਰਿਂ	ਲਿਂ	ਵਿਂ	ਸ਼ਿਂ	ਬਿਂ	ਸਿਂ	ਹਿਂ
ੀਂ	ਪੀਂ	ਫੀਂ	ਬੀਂ	ਭੀਂ	ਮੀਂ	ਯੀਂ	ਰੀਂ	ਲੀਂ	ਵੀਂ	ਸ਼ੀਂ	ਬੀਂ	ਸੀਂ	ਹੀਂ
ੁਂ	ਪੁਂ	ਫੁਂ	ਬੁਂ	ਭੁਂ	ਮੁਂ	ਯੁਂ	ਰੁਂ	ਲੁਂ	ਵੁਂ	ਸ਼ੁਂ	ਬੁਂ	ਸੁਂ	ਹੁਂ
ੂਂ	ਪੂਂ	ਫੂਂ	ਬੂਂ	ਭੂਂ	ਮੂਂ	ਯੂਂ	ਰੂਂ	ਲੂਂ	ਵੂਂ	ਸ਼ੂਂ	ਬੂਂ	ਸੂਂ	ਹੂਂ
ੈਂ	ਪੈਂ	ਫੈਂ	ਬੈਂ	ਭੈਂ	ਮੈਂ	ਯੈਂ	ਰੈਂ	ਲੈਂ	ਵੈਂ	ਸ਼ੈਂ	ਬੈਂ	ਸੈਂ	ਹੈਂ
ੋਂ ³⁵	ਪੋਂ	ਫੋਂ	ਬੋਂ	ਭੋਂ	ਮੋਂ	ਯੋਂ	ਰੋਂ	ਲੋਂ	ਵੋਂ	ਸ਼ੋਂ	ਬੋਂ	ਸੋਂ	ਹੋਂ
ੌਂ	ਪੌਂ	ਫੌਂ	ਬੌਂ	ਭੌਂ	ਮੌਂ	ਯੌਂ	ਰੌਂ	ਲੌਂ	ਵੌਂ	ਸ਼ੌਂ	ਬੌਂ	ਸੌਂ	ਹੌਂ
ੌਈਂ	ਪੌਈਂ	ਫੌਈਂ	ਬੌਈਂ	ਭੌਈਂ	ਮੌਈਂ	ਯੌਈਂ	ਰੌਈਂ	ਲੌਈਂ	ਵੌਈਂ	ਸ਼ੌਈਂ	ਬੈਈਂ	ਸੈਈਂ	ਹੈਈਂ
ਾਓਂ	ਪਾਓਂ	ਫਾਓਂ	ਬਾਓਂ	ਭਾਓਂ	ਮਾਓਂ	ਯਾਓਂ	ਰਾਓਂ	ਲਾਓਂ	ਵਾਓਂ	ਸਾਓਂ	ਬਾਓਂ	ਸਾਓਂ	ਹਾਓਂ
ਿਓਂ	ਪਿਓਂ	ਫਿਓਂ	ਬਿਓਂ	ਭਿਓਂ	ਮਿਓਂ	ਧਿਓਂ	ਰਿਓਂ	ਲਿਓਂ	ਵਿਓਂ	ਸ਼ਿਓਂ	ਬਿਓਂ	ਸਿਓਂ	ਹਿਓਂ
ੀਓਂ	ਪੀਓਂ	ਫੀਓਂ	ਬੀਓਂ	ਭੀਓਂ	ਮੀਓਂ	ਯੀਓਂ	ਰੀਓਂ	ਲੀਓਂ	ਵੀਓਂ	ਸ਼ੀਓਂ	ਬੀਓਂ	ਸੀਓਂ	ਹੀਓਂ

³⁵ Since ਝ, ਝੌ are used only for the purpose of writing loan words mainly from English. However, since such occurrences are rare, the nasal shapes of ਝੰ, ਝੌਂ 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	ज़	फ	ड़.	ठ़.
ं	कं	खं	गं	जं	फं	डं.	ठं.
ा	क्रा॑	ख्रा॑	గ्रा॑	জ্রা॑	ফ্রা॑	ড্রা॑	ঠ্রা॑
িং	ক্ৰি	খ্ৰি	গ্ৰি	জ্ৰি	ফ্ৰি	ড্ৰি	ঠ্ৰি
ীং	ক্ৰী	খ্ৰী	গ্ৰী	জ্ৰী	ফ্ৰী	ড্ৰী	ঠ্ৰী
ুং	ক্ৰু	খ্ৰু	গ্ৰু	জ্ৰু	ফ্ৰু	ড্ৰু	ঠ্ৰু
ুৰং	ক্ৰুৰ	খ্ৰুৰ	গ্ৰুৰ	জ্ৰুৰ	ফ্ৰুৰ	ড্ৰুৰ	ঠ্ৰুৰ
ুৰুং	ক্ৰুৰু	খ্ৰুৰু	গ্ৰুৰু	জ্ৰুৰু	ফ্ৰুৰু	ড্ৰুৰু	ঠ্ৰুৰু
ৈং	ক্ৰে	খ্ৰে	গ্ৰে	জ্ৰে	ফ্ৰে	ড্ৰে	ঠ্ৰে
োং ³⁶	ক্ৰো	খ্ৰো	গ্ৰো	জ্ৰো	ফ্ৰো	ড্ৰো	ঠ্ৰো
ৈোং	ক্ৰৈ	খ্ৰৈ	গ্ৰৈ	জ্ৰৈ	ফ্ৰৈ	ড্ৰৈ	ঠ্ৰৈ
োৰং	ক্ৰোৰ	খ্ৰোৰ	গ্ৰোৰ	জ্ৰোৰ	ফ্ৰোৰ	ড্ৰোৰ	ঠ্ৰোৰ
ৈোৰং	ক্ৰৈৰ	খ্ৰৈৰ	গ্ৰৈৰ	জ্ৰৈৰ	ফ্ৰৈৰ	ড্ৰৈৰ	ঠ্ৰৈৰ

³⁶ Since শ, চ are used only for the purpose of writing loan words mainly from English. However, since such occurrences are rare, the nasal shapes of শং, চং are rarely encountered in Hindi.

Consonant and Mātrā + Nasal combinations: With Candrabindu - Set 1

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.

This rule could create issues for OCR/OHWR if Chandrabindu is inputted over matras above the shirorekha and is automatically mapped to Anuswar as per the rule above. It is proposed that a Normalisation rule be introduced. In the case of OCR/OHWR which analyse the character at the display level, the combination matra+Anuswar will be shown as such.

	क	ख	ग	घ	ङ	च	छ	ज	झ	ञ
ं	ं	ং	ং	ং	ং	ং	ং	ং	ং	ং
াং	াং	াং	াং	াং	াং	াং	াং	াং	াং	াং
িং	িং	িং	িং	িং	িং	িং	িং	িং	িং	িং
ীং	ীং	াীং	াীং	াীং	াীং	াীং	াীং	াীং	াীং	াীং
ুং	ুং	াুং	াুং	াুং	াুং	াুং	াুং	াুং	াুং	াুং
ূং	ূং	াূং	াূং	াূং	াূং	াূং	াূং	াূং	াূং	াূং
ৃং	ৃং	াৃং	াৃং	াৃং	াৃং	াৃং	াৃং	াৃং	াৃং	াৃং
েং	েং	াং	াং	াং	াং	াং	াং	াং	াং	াং
ঁং	ঁং	াঁং	াঁং	াঁং	াঁং	াঁং	াঁং	াঁং	াঁং	াঁং
োং	োং	াোং	াোং	াোং	াোং	াোং	াোং	াোং	াোং	াোং
ুং	ুং	াুং	াুং	াুং	াুং	াুং	াুং	াুং	াুং	াুং
ৈং	ৈং	াৈং	াৈং	াৈং	াৈং	াৈং	াৈং	াৈং	াৈং	াৈং
ৌং	ৌং	াৌং	াৌং	াৌং	াৌং	াৌং	াৌং	াৌং	াৌং	াৌং

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. 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.

	ਟ	ਠ	ਛ	ਝ	ਯ	ਤ	ਥ	ਦ	ਧ	ਨ
ਾਂ	ਟਾਂ	ਠਾਂ	ਛਾਂ	ਝਾਂ	ਯਾਂ	ਤਾਂ	ਥਾਂ	ਦਾਂ	ਧਾਂ	ਨਾਂ
ਿਂ	ਟਿਂ	ਠਿਂ	ਛਿਂ	ਝਿਂ	ਯਿਂ	ਤਿਂ	ਥਿਂ	ਦਿਂ	ਧਿਂ	ਨਿਂ
ੀਂ	ਟੀਂ	ਠੀਂ	ਛੀਂ	ਝੀਂ	ਯੀਂ	ਤੀਂ	ਥੀਂ	ਦੀਂ	ਧੀਂ	ਨੀਂ
ੁਂ	ਟੁਂ	ਠੁਂ	ਛੁਂ	ਝੁਂ	ਯੁਂ	ਤੁਂ	ਥੁਂ	ਦੁਂ	ਧੁਂ	ਨੁਂ
੍ਰੂਂ	ਟ੍ਰੂਂ	ਠ੍ਰੂਂ	ਛ੍ਰੂਂ	ਝ੍ਰੂਂ	ਯ੍ਰੂਂ	ਤ੍ਰੂਂ	ਥ੍ਰੂਂ	ਦ੍ਰੂਂ	ਧ੍ਰੂਂ	ਨ੍ਰੂਂ
ੰ	ਟੰ	ਠੰ	ਛੰ	ਝੰ	ਯੰ	ਤੰ	ਥੰ	ਦੰ	ਧੰ	ਨੰ
ੈਂ	ਟੈਂ	ਠੈਂ	ਛੈਂ	ਝੈਂ	ਯੈਂ	ਤੈਂ	ਥੈਂ	ਦੈਂ	ਧੈਂ	ਨੈਂ
ੋਂ	ਟੋਂ	ਠੋਂ	ਛੋਂ	ਝੋਂ	ਯੋਂ	ਤੋਂ	ਥੋਂ	ਦੋਂ	ਧੋਂ	ਨੋਂ
੍ਰੋਂ ³⁷	ਟ੍ਰੋਂ	ਠ੍ਰੋਂ	ਛ੍ਰੋਂ	ਝ੍ਰੋਂ	ਯ੍ਰੋਂ	ਤ੍ਰੋਂ	ਥ੍ਰੋਂ	ਦ੍ਰੋਂ	ਧ੍ਰੋਂ	ਨ੍ਰੋਂ
ੌਂ	ਟੌਂ	ਠੌਂ	ਛੌਂ	ਝੌਂ	ਯੌਂ	ਤੌਂ	ਥੌਂ	ਦੌਂ	ਧੌਂ	ਨੌਂ

³⁷ Since ଶ, ଶ୍ୟ are used only for the purpose of writing loan words mainly from English. However, since such occurrences are rare, the nasal shapes of ଶନ୍, ଶ୍ୟନ୍ 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. 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.

	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	रृँ	लृँ	वृँ	शृँ	षृँ	सृँ	हृँ
ेँ	पेँ	फेँ	बेँ	ভেঁ	মেঁ	যেঁ	রেঁ	লেঁ	বেঁ	শেঁ	ষেঁ	সেঁ	হেঁ
ঁ	পঁ	ফঁ	বঁ	ভঁ	মঁ	যঁ	রঁ	লঁ	বঁ	শঁ	ষঁ	সঁ	হঁ
ৈঁ	পৈঁ	ফৈঁ	বৈঁ	ভৈঁ	মৈঁ	যৈঁ	ৈঁ	লৈঁ	বৈঁ	শৈঁ	ষৈঁ	সৈঁ	হৈঁ
োঁ	পোঁ	ফোঁ	বোঁ	ভোঁ	মোঁ	যোঁ	োঁ	লোঁ	বোঁ	শোঁ	ষোঁ	সোঁ	হোঁ
াুঁ	পাঁ	ফাঁ	বাঁ	ভাঁ	মাঁ	যাঁ	াুঁ	লাঁ	বাঁ	শাঁ	ষাঁ	সাঁ	হাঁ
ৌঁ	পৌঁ	ফৌঁ	বৌঁ	ভৌঁ	মৌঁ	যৌঁ	ৌঁ	লৌঁ	বৌঁ	শৌঁ	ষৌঁ	সৌঁ	হৌঁ

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. 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.

	कः	खः	गः	जः	फः	छः	ठः
ॐ	क्रँ	ख्यँ	गँ	जँ	फँ	छँ	ठँ
ा०	क्राँ	ख्याँ	गा०	जा०	फ्राँ	छा०	ठा०
ि०	क्रि	ख्यि	गि	जि	फ्रि	छि	ठि०
ी०	क्री	ख्यी	गी०	जी०	फ्री०	छी०	ठी०
ु०	क्रु०	ख्यु०	गु०	जु०	फ्रु०	छु०	ठु०
ू०	क्रू०	ख्यू०	गू०	जू०	फ्रू०	छू०	ठू०
ृ०	क्रृ०	ख्यृ०	गृ०	जृ०	फ्रृ०	छृ०	ठृ०
े०	क्रे०	ख्ये०	गे०	जे०	फ्रे०	छे०	ठे०
ो०	क्रो०	ख्यो०	गो०	जो०	फ्रो०	छो०	ठो०
ौ०	क्रौ०	ख्यौ०	गौ०	जौ०	फ्रौ०	छौ०	ठौ०
ौ०	क्रौ०	ख्यौ०	गौ०	जौ०	फ्रौ०	छौ०	ठौ०

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	एँ	ए	ऐ	ओँ	ओ	औ
ँ	अँ	आँ	इँ	ईँ	उँ	ऊँ	ऋँ	एँ	एँ	ऐँ	ओँ	ओँ	औँ
ं	अं	आं	इं	ईं	उं	ऊं	ऋं	एं	एं	ऐं	ओं	ओं	औं

Consonant and Mātrā combinations for Main Ligatures क्ष, ज्ञ, त्र

Although क्ष, ज्ञ, त्र and श्र are not consonants per se³⁸, it was felt that for the font developer all matra combinations with these three ligatural shapes be provided for the font developer.

Set 1 Consonant and Mātrā combinations

	क्ष	ज्ञ	त्र	श्र
ा	क्षा	ज्ञा	त्रा	श्रा
ि	क्षि	ज्ञि	त्रि	श्रि
ी	क्षी	ज्ञी	त्री	श्री
ु	क्षु	ज्ञु	त्रु	श्रु
ू	क्षू	ज्ञू	त्रू	श्रू
ृ	क्षृ	ज्ञृ	त्रृ	श्रृ
े	क्षे	ज्ञे	त्रे	श्रे
ঁ	ক্ষঁ	জ্ঞঁ	ত্রঁ	শ্রঁ
ৈ	ক্ষৈ	জ্ঞৈ	ত্রৈ	শ্রৈ
ো	ক্ষো	জ্ঞো	ত্রো	শ্রো
ুঁ	ক্ষুঁ	জ্ঞুঁ	ত্রুঁ	শ্রুঁ
ৌ	ক্ষৌ	জ্ঞৌ	ত্রৌ	শ্রৌ

³⁸ Cf. p. 22

Set 2 Consonant and Mātrā + Nasal combinations: with Anuswara

	क्ष	ज्ञ	त्र	श्र
ं	क्षं	ज्ञं	त्रं	श्रं
ां	क्षां	ज्ञां	त्रां	श्रां
िं	क्षिं	ज्ञिं	त्रिं	श्रिं
ीं	क्षीं	ज्ञीं	त्रीं	श्रीं
ूं	क्षुं	ज्ञुं	त्रुं	श्रुं
ूूं	क्षूूं	ज्ञूूं	त्रूूं	श्रूूं
ृं	क्षृं	ज्ञृं	त्रृं	श्रृं
ें	क्षें	ज्ञें	त्रें	श्रें
ँ	क्षँ	ज्ञँ	त्रँ	श्रँ
ैं	क्षैं	ज्ञैं	त्रैं	श्रैं
ों	क्षों	ज्ञों	त्रों	श्रों
ॉं	क्षॉं	ज्ञॉं	त्रॉं	श्रॉं
ौं	क्षौं	ज्ञौं	त्रौं	श्रौं

Set 3 Consonant and Mātrā + Nasal combinations: With Candrabindu

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.

	क्ष	ज्ञ	त्र	श्र
ं	क्षं	ज्ञं	त्रं	श्रं
ां	क्षां	ज्ञां	त्रां	श्रां
िं	क्षिं	ज्ञिं	त्रिं	श्रिं
ीं	क्षीं	ज्ञीं	त्रीं	श्रीं
ूं	क्षुं	ज्ञुं	त्रुं	श्रुं
ूूं	क्षूूं	ज्ञूूं	त्रूूं	श्रूूं
ृं	क्षृं	ज्ञृं	त्रृं	श्रृं
ें	क्षें	ज्ञें	त्रें	श्रें
ँ	क्षँ	ज्ञँ	त्रँ	श्रँ
ैं	क्षैं	ज्ञैं	त्रैं	श्रैं
ों	क्षों	ज्ञों	त्रों	श्रों
ॉं	क्षॉं	ज्ञॉं	त्रॉं	श्रॉं
ौं	क्षौं	ज्ञौं	त्रौं	श्रौं

ଓেঁ	କ୍ଷାଁ	ଜ୍ଞାଁ	ତ୍ରାଁ	ଶ୍ରାଁ
ଠୋଁ	କ୍ଷାଁ	ଜ୍ଞାଁ	ତ୍ରାଁ	ଶ୍ରାଁ
ଠୌଁ	କ୍ଷାଁ	ଜ୍ଞାଁ	ତ୍ରାଁ	ଶ୍ରାଁ

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”³⁹.

CHC(combination of two consonants) - Set 1										
	क	ख	ग	घ	ङ	च	छ	ज	झ	ञ
क्	क्क ⁴⁰	क्ख	क्ग	क्घ	क्ङ	क्च	क्छ	क्ज	क्झ	क्ञ
ख्	ख्क	ख्ख	ख्ग	ख्घ	ख्ङ	ख्च	ख्छ	ख्ज	ख्झ	ख्ञ
ग्	ग्क	ग्ख	ग्ग	ग्घ	ग्ङ	ग्च	ग्छ	ग्ज	ग्झ	ग्ञ
घ्	घ्क	घ्ख	घ्ग	घ्घ	घ्ङ	घ्च	घ्छ	घ्ज	घ्झ	घ्ञ
ङ्	ङ्क	ङ्ख	ङ्ग	ङ्घ	ङ्ङ	ङ्च	ङ्छ	ङ्ज	ङ्झ	ङ्ञ
च्	च्क	च्ख	च्ग	च्घ	च्ङ	च्च	च्छ	च्ज	च्झ	च्ञ
छ्	छ्क	छ्ख	छ्ग	छ्घ	छ्ङ	छ्च	छ्छ	छ्ज	छ्झ	छ्ञ
ज्	ज्क	ज्ख	ज्ग	ज्घ	ज्ङ	ज्च	ज्छ	ज्ज	ज्झ	ज्ञ
झ्	झ्क	झ्ख	झ्ग	झ्घ	झ्ङ	झ्च	झ्छ	झ्ज	झ्झ	झ्ञ
ञ्	ञ्क	ञ्ख	ञ्ग	ञ्घ	ञ्ङ	ञ्च	ञ्छ	ञ्ज	ञ्झ	ञ्ञ
ट्	ट्क	ट्ख	ट्ग	ट्घ	ट्ङ	ट्च	ट्छ	ट्ज	ट्झ	ट्ञ

³⁹ ZWJ/ZWNJ are generally not used in Hindi. cf. Annexure 3: Note on ZWJ/ZWNJ in Devanāgarī.

⁴⁰ Although the Central Hindi Directorate prefers the linear form of the conjunct, the use of ं i.e. a stacked form is also permissible and is therefore noted as such.

ਠ	ਠਕ	ਠਖ	ਠਗ	ਠਘ	ਠਡ	ਠਚ	ਠਛ	ਠਜ	ਠੜ	ਠਯ
ਡ	ਡਕ	ਡਖ	ਡਗ	ਡਘ	ਡਡ	ਡਚ	ਡਛ	ਡਜ	ਡੜ	ਡਯ
ਛ	ਛਕ	ਛਖ	ਛਗ	ਛਘ	ਛਡ	ਛਚ	ਛਛ	ਛਜ	ਛੜ	ਛਯ
ਣ	ਣਕ	ਣਖ	ਣਗ	ਣਘ	ਣਡ	ਣਚ	ਣਛ	ਣਜ	ਣੜ	ਣਯ
ਤ	ਤਕ	ਤਖ	ਤਗ	ਤਘ	ਤਡ	ਤਚ	ਤਛ	ਤਜ	ਤੜ	ਤਯ
ਥ	ਥਕ	ਥਖ	ਥਗ	ਥਘ	ਥਡ	ਥਚ	ਥਛ	ਥਜ	ਥੜ	ਥਯ
ਦ	ਦਕ	ਦਖ	ਦਗ	ਦਘ	ਦਡ	ਦਚ	ਦਛ	ਦਜ	ਦੜ	ਦਯ
ਧ	ਧਕ	ਧਖ	ਧਗ	ਧਘ	ਧਡ	ਧਚ	ਧਛ	ਧਜ	ਧੜ	ਧਯ
ਨ	ਨਕ	ਨਖ	ਨਗ	ਨਘ	ਨਡ	ਨਚ	ਨਛ	ਨਜ	ਨੜ	ਨਯ
ਪ	ਪਕ	ਪਖ	ਪਗ	ਪਘ	ਪਡ	ਪਚ	ਪਛ	ਪਜ	ਪੜ	ਪਯ
ਫ	ਫਕ	ਫਖ	ਫਗ	ਫਘ	ਫਡ	ਫਚ	ਫਛ	ਫਜ	ਫੜ	ਫਯ
ਬ	ਬਕ	ਬਖ	ਬਗ	ਬਘ	ਬਡ	ਬਚ	ਬਛ	ਬਜ	ਬੜ	ਬਯ
ਭ	ਭਕ	ਭਖ	ਭਗ	ਭਘ	ਭਡ	ਭਚ	ਭਛ	ਭਜ	ਭੜ	ਭਯ
ਮ	ਮਕ	ਮਖ	ਮਗ	ਮਘ	ਮਡ	ਮਚ	ਮਛ	ਮਜ	ਮੜ	ਮਯ
ਧ	ਧਕ	ਧਖ	ਧਗ	ਧਘ	ਧਡ	ਧਚ	ਧਛ	ਧਜ	ਧੜ	ਧਯ
ਰ	ਰਕ	ਰਖ	ਰਗ	ਰਘ	ਰਡ	ਰਚ	ਰਛ	ਰਜ	ਰੜ	ਰਯ
ਲ	ਲਕ	ਲਖ	ਲਗ	ਲਘ	ਲਡ	ਲਚ	ਲਛ	ਲਜ	ਲੜ	ਲਯ
ਵ	ਵਕ	ਵਖ	ਵਗ	ਵਘ	ਵਡ	ਵਚ	ਵਛ	ਵਜ	ਵੜ	ਵਯ
ਸ਼	ਸ਼ਕ	ਸ਼ਖ	ਸ਼ਗ	ਸ਼ਘ	ਸ਼ਡ	ਸ਼ਚ	ਸ਼ਛ	ਸ਼ਜ	ਸ਼ੜ	ਸ਼ਯ
਷	਷ਕ	਷ਖ	਷ਗ	਷ਘ	਷ਡ	਷ਚ	਷ਛ	਷ਜ	਷ੜ	਷ਯ
ਸ	ਸਕ	ਸਖ	ਸਗ	ਸਘ	ਸਡ	ਸਚ	ਸਛ	ਸਜ	ਸੜ	ਸਯ
ਹ	ਹਕ	ਹਖ	ਹਗ	ਹਘ	ਹਡ	ਹਚ	ਹਛ	ਹਜ	ਹੜ	ਹਯ
ਕੁ	ਕੁਕ	ਕੁਖ	ਕੁਗ	ਕੁਘ	ਕੁਡ	ਕੁਚ	ਕੁਛ	ਕੁਜ	ਕੁੜ	ਕੁਯ
ਖੁ	ਖੁਕ	ਖੁਖ	ਖੁਗ	ਖੁਘ	ਖੁਡ	ਖੁਚ	ਖੁਛ	ਖੁਜ	ਖੁੜ	ਖੁਯ
ਗੁ	ਗੁਕ	ਗੁਖ	ਗੁਗ	ਗੁਘ	ਗੁਡ	ਗੁਚ	ਗੁਛ	ਗੁਜ	ਗੁੜ	ਗੁਯ
ਜੁ	ਜੁਕ	ਜੁਖ	ਜੁਗ	ਜੁਘ	ਜੁਡ	ਜੁਚ	ਜੁਛ	ਜੁਜ	ਜੁੜ	ਜੁਯ
ਫੁ	ਫੁਕ	ਫੁਖ	ਫੁਗ	ਫੁਘ	ਫੁਡ	ਫੁਚ	ਫੁਛ	ਫੁਜ	ਫੁੜ	ਫੁਯ
ਡੁ	ਡੁਕ	ਡੁਖ	ਡੁਗ	ਡੁਘ	ਡੁਡ	ਡੁਚ	ਡੁਛ	ਡੁਜ	ਡੁੜ	ਡੁਯ
ਛੁ	ਛੁਕ	ਛੁਖ	ਛੁਗ	ਛੁਘ	ਛੁਡ	ਛੁਚ	ਛੁਛ	ਛੁਜ	ਛੁੜ	ਛੁਯ

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 “कट”.

CHC(combination of two consonants) - Set 2											
	ट	ठ	ड	ढ	ण	त	थ	द	ध	न	
क्	कट	कठ	कड	कढ	कण	कत	कथ	कद	कध	कन	
ख्	खट	खठ	खड	खढ	खण	खत	खथ	खद	खধ	খন	
গ্	গট	গঠ	গড	গঢ	গণ	গত	গথ	গদ	গধ	গন	
ঘ্	ঘট	ঘঠ	ঘড	ঘঢ	ঘণ	ঘত	ঘথ	ঘদ	ঘধ	ঘন	
ঙ্	ঙ্ট	ঙ্ঠ	ঙ্ড	ঙ্ঢ	ঙ্ণ	ঙ্ত	ঙ্থ	ঙ্দ	ঙ্ধ	ঙ্ন	
চ্	চট	চঠ	চড	চঢ	চণ	চত	চথ	চদ	চধ	চন	
ছ্	ছট	ছঠ	ছড	ছঢ	ছণ	ছত	ছথ	ছদ	ছধ	ছন	
জ্	জট	জঠ	জড	জঢ	জণ	জত	জথ	জদ	জধ	জন	
ঝ্	ঝট	ঝঠ	ঝড	ঝঢ	ঝণ	ঝত	ঝথ	ঝদ	ঝধ	ঝন	
ঝ্	ঝট	ঝঠ	ঝড	ঝঢ	ঝণ	ঝত	ঝথ	ঝদ	ঝধ	ঝন	
ঝ্	ঝট	ঝঠ	ঝড	ঝঢ	ঝণ	ঝত	ঝথ	ঝদ	ঝধ	ঝন	
ঠ্	ঠট	ঠঠ	ঠড	ঠঢ	ঠণ	ঠত	ঠথ	ঠদ	ঠধ	ঠন	
ঠ্	ঠট	ঠঠ	ঠড	ঠঢ	ঠণ	ঠত	ঠথ	ঠদ	ঠধ	ঠন	
ঙ্	ঙট	ঙঠ	ঙড	ঙঢ	ঙণ	ঙত	ঙথ	ঙদ	ঙধ	ঙন	
ঙ্	ঙট	ঙঠ	ঙড	ঙঢ	ঙণ	ঙত	ঙথ	ঙদ	ঙধ	ঙন	
ঠ্	ঠট	ঠঠ	ঠড	ঠঢ	ঠণ	ঠত	ঠথ	ঠদ	ঠধ	ঠন	
ঠ্	ঠট	ঠঠ	ঠড	ঠঢ	ঠণ	ঠত	ঠথ	ঠদ	ঠধ	ঠন	
ণ্	ণট	ণঠ	ণড	ণঢ	ণণ	ণত	ণথ	ণদ	ণধ	ণন	
ণ্	ণট	ণঠ	ণড	ণঢ	ণণ	ণত	ণথ	ণদ	ণধ	ণন	
ত্	তট	তঠ	তড	তঢ	তণ	তত	তথ	তদ	তধ	তন	
ত্	তট	তঠ	তড	তঢ	তণ	তত	তথ	তদ	তধ	তন	
থ্	থট	থঠ	থড	থঢ	থণ	থত	থথ	থদ	থধ	থন	
থ্	থট	থঠ	থড	থঢ	থণ	থত	থথ	থদ	থধ	থন	
দ্	দট	দঠ	দড	দঢ	দণ	দত	দথ	দদ	দধ	দন	
ধ্	ধট	ধঠ	ধড	ধঢ	ধণ	ধত	ধথ	ধদ	ধধ	ধন	
ন্	নট	নঠ	নড	নঢ	নণ	নত	নথ	নদ	নধ	নন	
প্	পট	পঠ	পড	পঢ	পণ	পত	পথ	পদ	পধ	পন	
ফ্	ফট	ফঠ	ফড	ফঢ	ফণ	ফত	ফথ	ফদ	ফধ	ফন	

ਾ	ਾਟ	ਾਲ	ਾਲੁ	ਾਲੁ	ਾਣ	ਾਣ	ਾਤ	ਾਥ	ਾਲੁ	ਾਲੁ	ਾਵ	ਾਨ
ਾਭ	ਾਭਟ	ਾਭਰ	ਾਭੁ	ਾਭੁ	ਾਭਣ	ਾਭਣ	ਾਭਤ	ਾਭਥ	ਾਭੁ	ਾਭੁ	ਾਭ	ਾਨ
ਾਮ	ਾਮਟ	ਾਮਰ	ਾਮੁ	ਾਮੁ	ਾਮਣ	ਾਮਣ	ਾਮਤ	ਾਮਥ	ਾਮੁ	ਾਮੁ	ਾਮ	ਾਨ
ਾਧ	ਾਧਟ	ਾਧਰ	ਾਧੁ	ਾਧੁ	ਾਧਾ	ਾਧਾ	ਾਧਤ	ਾਧਥ	ਾਧੁ	ਾਧੁ	ਾਧ	ਾਨ
ਾਰ	ਾਰਟ	ਾਰਿ	ਾਰੁ	ਾਰੁ	ਾਰਣ	ਾਰਣ	ਾਰਤ	ਾਰਥ	ਾਰੁ	ਾਰੁ	ਾਰ	ਾਨ
ਾਲ	ਾਲਟ	ਾਲਰ	ਾਲੁ	ਾਲੁ	ਾਲਣ	ਾਲਣ	ਾਲਤ	ਾਲਥ	ਾਲੁ	ਾਲੁ	ਾਲ	ਾਨ
ਾਵ	ਾਵਟ	ਾਵਿ	ਾਵੁ	ਾਵੁ	ਾਵਾ	ਾਵਾ	ਾਵਤ	ਾਵਥ	ਾਵੁ	ਾਵੁ	ਾਵ	ਾਨ
ਾਸ	ਾਸਟ	ਾਸਰ	ਾਸੁ	ਾਸੁ	ਾਸਣ	ਾਸਣ	ਾਸਤ	ਾਸਥ	ਾਸੁ	ਾਸੁ	ਾਸ	ਾਨ
ਾ਷	ਾ਷ਟ	ਾ਷ਿ	ਾ਷ੁ	ਾ਷ੁ	ਾ਷ਾ	ਾ਷ਾ	ਾ਷ਤ	ਾ਷ਥ	ਾ਷ੁ	ਾ਷ੁ	ਾ਷	ਾਨ
ਾਸ	ਾਸਟ	ਾਸਰ	ਾਸੁ	ਾਸੁ	ਾਸਣ	ਾਸਣ	ਾਸਤ	ਾਸਥ	ਾਸੁ	ਾਸੁ	ਾਸ	ਾਨ
ਾਹ	ਾਹਟ	ਾਹਰ	ਾਹੁ	ਾਹੁ	ਾਹਣ	ਾਹਣ	ਾਹਤ	ਾਹਥ	ਾਹੁ	ਾਹੁ	ਾਹ	ਾਨ
ਾਕ	ਾਕਟ	ਾਕਰ	ਾਕੁ	ਾਕੁ	ਾਕਣ	ਾਕਣ	ਾਕਤ	ਾਕਥ	ਾਕੁ	ਾਕੁ	ਾਕ	ਾਨ
ਾਖ	ਾਖਟ	ਾਖਰ	ਾਖੁ	ਾਖੁ	ਾਖਣ	ਾਖਣ	ਾਖਤ	ਾਖਥ	ਾਖੁ	ਾਖੁ	ਾਖ	ਾਨ
ਾਗ	ਾਗਟ	ਾਗਰ	ਾਗੁ	ਾਗੁ	ਾਗਣ	ਾਗਣ	ਾਗਤ	ਾਗਥ	ਾਗੁ	ਾਗੁ	ਾਗ	ਾਨ
ਾਜ	ਾਜਟ	ਾਜਰ	ਾਜੁ	ਾਜੁ	ਾਜਾ	ਾਜਾ	ਾਜਤ	ਾਜਥ	ਾਜੁ	ਾਜੁ	ਾਜ	ਾਨ
ਾਫ	ਾਫਟ	ਾਫਰ	ਾਫੁ	ਾਫੁ	ਾਫਣ	ਾਫਣ	ਾਫਤ	ਾਫਥ	ਾਫੁ	ਾਫੁ	ਾਫ	ਾਨ
ਾਡ	ਾਡਟ	ਾਡਰ	ਾਡੁ	ਾਡੁ	ਾਡਣ	ਾਡਣ	ਾਡਤ	ਾਡਥ	ਾਡੁ	ਾਡੁ	ਾਡ	ਾਨ

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	क्षि	क्स	क्ह	
ख्	ख्प	ख्फ	ख्ब	ख्भ	ख्म	ख्य	ख्र	ख्ल	ख्व	ख्श	ख्षि	ख्स	ख्ह	
ग्	ग्प	ग्फ	ग्ब	ग्भ	ग्म	ग्य	ग्र	ग्ल	ग्व	ग्श	ग्षि	ग्स	ग्ह	
घ्	घ्प	घ्फ	घ्ब	घ्भ	घ्म	घ्य	घ्र	घ्ल	घ्व	घ्श	घ्षि	घ्स	घ्ह	
ङ्	ङ्प	ङ्फ	ङ्ब	ङ्भ	ङ्म	ङ्य	ङ्र	ङ्ल	ङ्व	ङ्श	ङ्षि	ङ्स	ङ्ह	
च्	च्प	च्फ	च्ब	च्भ	च्म	च्य	च्र	च्ल	च्व	च्श	च्षि	च्स	च्ह	
छ्	छ्प	छ्फ	छ्ब	छ्भ	छ्म	छ्य	छ्र	छ्ल	छ्व	छ्श	छ्षि	छ्स	छ्ह	
ज्	ज्प	ज्फ	ज्ब	ज्भ	ज्म	ज्य	ज्र	ज्ल	ज्व	ज्श	ज्षि	ज्स	ज्ह	
झ्	झ्प	झ्फ	झ्ब	झ्भ	झ्म	झ्य	झ्र	झ्ल	झ्व	झ्श	झ्षि	झ्स	झ्ह	
ञ्	ञ्प	ञ्फ	ञ्ब	ञ्भ	ञ्म	ञ्य	ञ्र	ञ्ल	ञ्व	ञ्श	ञ्षि	ञ्स	ञ्ह	
ट्	ट्प	ट्फ	ट्ब	ट्भ	ट्म	ट्य	ट्र	ट्ल	ट्व	ट्श	ट्षि	ट्स	ट्ह	
ठ्	ठ्प	ठ्फ	ठ्ब	ठ्भ	ठ्म	ठ्य	ठ्र	ठ्ल	ठ्व	ठ्श	ठ्षि	ठ्स	ठ्ह	
ड्	ड्प	ड्फ	ड्ब	ड्भ	ड्म	ड्य	ड्र	ड्ल	ड्व	ड्श	ड्षि	ड्स	ड्ह	
ढ्	ढ्प	ढ्फ	ढ्ब	ढ्भ	ढ्म	ढ्य	ढ्र	ढ्ल	ढ्व	ढ्श	ढ्षि	ढ्स	ढ्ह	
ण्	ण्प	ण्फ	ण्ब	ण्भ	ण्म	ण्य	ण्र	ण्ल	ण्व	ण्श	ण्षि	ण्स	ण्ह	
त्	त्प	त्फ	त्ब	त्भ	त्म	त्य	त्र	त्ल	त्व	त्श	त्षि	त्स	त्ह	
थ्	थ्प	थ्फ	थ्ब	थ्भ	थ्म	थ्य	थ्र	थ्ल	थ्व	थ्श	थ्षि	थ्स	थ्ह	
द्	द्प	द्फ	द्ब	द्भ	द्म	द्य	द्र	द्ल	द्व	द्श	द्षि	द्स	द्ह	
ध्	ध्प	ध्फ	ध्ब	ध्भ	ধ্ম	ধ্য	ধ্র	ধ্ল	ধ্঵	ধ্শ	ধ্ষি	ধ্স	ধ্হ	
ন्	ন्प	ন্ফ	ন্ব	ন্ভ	ন্ম	ন্য	ন্র	ন্ল	ন্঵	ন্শ	ন্ষি	ন্স	ন্হ	
প্	প্প	প্ফ	প্ব	প্ভ	প্ম	প্য	প্র	প্ল	প্঵	প্শ	প্ষি	প্স	প্হ	
ফ্	ফ্প	ফ্ফ	ফ্ব	ফ্ভ	ফ্ম	ফ্য	ফ্র	ফ্ল	ফ্঵	ফ্শ	ফ্ষি	ফ্স	ফ্হ	

⁴¹ An important combination very often mistyped is श+ङ् = श् which is often written as श+र+ङ् = श्. 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 “**କ**” and “**ଙ୍କ**” is ligature “**କ୍ଙ୍କ**”.

CHC Set 4								
	ਕੁ	ਖੁ	ਗੁ	ਜੁ	ਫੁ	ਡੁ	ਢੁ	
ਕ	ਕਕ	ਕਖ	ਕਾ	ਕਯ	ਕਫ	ਕਡ	ਕਢ	
ਖ	ਖਕ	ਖਖ	ਖਾ	ਖਯ	ਖਫ	ਖਡ	ਖਢ	
ਗ	ਗਕ	ਗਖ	ਗਾ	ਗਯ	ਗਫ	ਗਡ	ਗਢ	
ਧ	ਧਕ	ਧਖ	ਧਾ	ਧਯ	ਧਫ	ਧਡ	ਧਢ	
ਙ	ਙਕ	ਙਖ	ਙਾ	ਙਯ	ਙਫ	ਙਡ	ਙਢ	
ਚ	ਚਕ	ਚਖ	ਚਾ	ਚਯ	ਚਫ	ਚਡ	ਚਢ	
ਛ	ਛਕ	ਛਖ	ਛਾ	ਛਯ	ਛਫ	ਛਡ	ਛਢ	
ਜ	ਜਕ	ਜਖ	ਜਾ	ਜਯ	ਜਫ	ਜਡ	ਜਢ	
ਝ	ਝਕ	ਝਖ	ਝਾ	ਝਯ	ਝਫ	ਝਡ	ਝਢ	
ਤ	ਤਕ	ਤਖ	ਤਾ	ਤਯ	ਤਫ	ਤਡ	ਤਢ	
ਠ	ਠਕ	ਠਖ	ਠਾ	ਠਯ	ਠਫ	ਠਡ	ਠਢ	
ਡ	ਡਕ	ਡਖ	ਡਾ	ਡਯ	ਡਫ	ਡਡ	ਡਢ	
ਣ	ਣਕ	ਣਖ	ਣਾ	ਣਯ	ਣਫ	ਣਡ	ਣਢ	
ਤ	ਤਕ	ਤਖ	ਤਾ	ਤਯ	ਤਫ	ਤਡ	ਤਢ	
ਥ	ਥਕ	ਥਖ	ਥਾ	ਥਯ	ਥਫ	ਥਡ	ਥਢ	
ਦ	ਦਕ	ਦਖ	ਦਾ	ਦਯ	ਦਫ	ਦਡ	ਦਢ	
ਧ	ਧਕ	ਧਖ	ਧਾ	ਧਯ	ਧਫ	ਧਡ	ਧਢ	
ਨ	ਨਕ	ਨਖ	ਨਾ	ਨਯ	ਨਫ	ਨਡ	ਨਢ	
ਪ	ਪਕ	ਪਖ	ਪਾ	ਪਯ	ਪਫ	ਪਡ	ਪਢ	
ਫ	ਫਕ	ਫਖ	ਫਾ	ਫਯ	ਫਫ	ਫਡ	ਫਢ	

ਅ	ਅਕ	ਅਖ	ਅਾ	ਅਜ	ਅਫ	ਅਡ	ਅਲ
ਭ	ਭਕ	ਭਖ	ਭਾ	ਭਜ	ਭਫ	ਭਡ	ਭਲ
ਮ	ਮਕ	ਮਖ	ਮਾ	ਮਜ	ਮਫ	ਮਡ	ਮਲ
ਧ	ਧਕ	ਧਖ	ਧਾ	ਧਜ	ਧਫ	ਧਡ	ਧਲ
ਰ	ਰਕ	ਰਖ	ਰਾ	ਰਜ	ਰਫ	ਰਡ	ਰਲ
ਲ	ਲਕ	ਲਖ	ਲਾ	ਲਜ	ਲਫ	ਲਡ	ਲਲ
ਵ	ਵਕ	ਵਖ	ਵਾ	ਵਜ	ਵਫ	ਵਡ	ਵਲ
ਸ	ਸਕ	ਸਖ	ਸਾ	ਸਜ	ਸਫ	ਸਡ	ਸਲ
ਹ	ਹਕ	ਹਖ	ਹਾ	ਹਜ	ਹਫ	ਹਡ	ਹਲ
ਕ	ਕਕ	ਕਖ	ਕਾ	ਕਜ	ਕਫ	ਕਡ	ਕਲ
ਖ	ਖਕ	ਖਖ	ਖਾ	ਖਜ	ਖਫ	ਖਡ	ਖਲ
ਗ	ਗਕ	ਗਖ	ਗਾ	ਗਜ	ਗਫ	ਗਡ	ਗਲ
ਜ	ਜਕ	ਜਖ	ਜਾ	ਜਜ	ਜਫ	ਜਡ	ਜਲ
ਫ	ਫਕ	ਫਖ	ਫਾ	ਫਜ	ਫਫ	ਫਡ	ਫਲ
ਡ	ਡਕ	ਡਖ	ਡਾ	ਡਜ	ਡਫ	ਡਡ	ਡਲ
ਲੁ	ਲੁਕ	ਲੁਖ	ਲੁਗ	ਲੁਜ	ਲੁਫ	ਲੁਡ	ਲੁਲ

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.⁴²

⁴² Quite a few examples are culled from Bahri and Nalanda Dictionaries. Others are taken form the CIIL corpus.

CHCHC	Example
कक्ष	दृक्क्षेप
कट्र	इलेक्ट्रोन
कट्स	स्पैक्ट्रस
कत्य	युक्त्याभास
कत्र	पंचवक्त्र
कत्व	पृथक्त्व
कत्र	वाक्प्रतोद
कत्य	शौकल्य
क्षण	तीक्ष्ण
क्षम	भाग्यलक्ष्मी
क्ष्य	अनपेक्ष्य
क्ष्य	इक्ष्याकु
कर्स्ड	फिक्स्ड
कर्स्त	वाक्स्तंभ
कर्स्व	वाक्स्वातंत्र्य
रुल्प	रुल्पाजा
ग्र	वणिग्राम
गज्य	प्राग्ज्योतिष
गद्ध	प्राग्द्वार
गन्य	अग्न्याधान
गभ्र	दिग्भ्रम
गव्य	वाग्व्यवहार
च्छ्व	उच्छ्वास
ज्ज्य	उज्ज्यल

त्वल	उत्वलेश
त्क्ष	तात्क्षणिक
त्थ	सुप्तोत्थित
त्व	तत्त्व
त्र	उत्प्रेरक
त्प्ल	उत्प्लव
त्य	स्थानमहात्म्य
त्र्य	वैचित्र्य
त्स्थ	तत्स्थानीय
त्स्न	जोत्स्ना
त्प्प	ईषत्पृष्ठ
त्प्प	हृत्पंदों
त्स्फ	उत्स्फूर्त
त्स्य	मत्स्य
त्स्व	तत्स्वरूप
द्‌ग्र	उद्‌ग्रीव
द्‌श	सद्‌ज्ञान
द्‌द्व	एतद्‌द्वारा
द्‌दृष	नामोदि‌दृष्ट
द्‌भ्र	उद्‌भ्रम
द्‌व्य	दारिद्र्य
द्‌व्य	सद्‌व्यवहारों
न्व	इन्वायरी
न्त्र	इन्फैन्ट्रीयाँ
न्ट्स	स्टूडन्ट्स

न्त्य	अचिन्त्य
न्त्र	उप-मन्त्री
न्द्य	अनिन्द्य
न्द्र	अंतरिक्षकेन्द्र
न्द्र	उपेन्द्रनाथ
न्ध्य	फलवन्ध्य
न्य	वैभिन्न्य
न्फ	कान्फ्रेस
प्त्य	प्राप्त्याशा
फ्ट्स	एयरक्राफ्ट्स
ब्ज्य	कौब्ज्य ⁴³
म्म्र	सम्माज
क्र्य	सकर्यूलर
क्र्क	अंतर्क्रिया
कर्स	माकर्सवाद
ख्य	मौख्य
ग्य	वग्य
ग्र	अंतर्ग्रस्त
घ्य	अघ्य
च्छ	मूर्च्छित
च्य	अच्य
ज्ञ	अंतज्ञान
ज्य	वज्य
ट्र्य	पोट्रेट

ट्र्स	स्पोटर्समैन
ण्य	वण्य
त्य	अतिमत्य
त्र्व	सात्र्व
त्स	वत्स
थ्र्य	सामथ्र्यो
झ्ड	आध्ड
र्य	सौहार्य
ट्र्द	आर्द्र
द्व्य	अंतद्व्यद्व ⁴⁴
ध्द	पूर्वाध्द
ध्य	अंतध्यान
ध्व	उध्व
प्र्य	समप्र्य
प्र्व	अंतप्रादेशिक
फ्य	कफ्यू
भ्र	निभ्राति
म्य	हम्य
ल्ड	फ्रीवल्ड
ल्स	चाल्स
व्य	बहिर्व्यसनी
व्र्व	निव्रण
श्व	पाश्व

⁴³ Sanskrit loan.: A hump-backed bull cf.

⁴⁴ अंतद्व्यद्व . मनस्ताप - mental agony Arvind Lexicon.

र्ष	कार्ष्ण
र्ष्य	ईर्ष्या
स्ट	फ़स्ट
ल्व्य	याज्ञवल्व्य
ल्ट्र	पोल्ट्रीफार्म
ल्ड्र	चिल्ड्रन
ल्प्य	प्रकल्प्य
ल्भ्य	प्रागल्भ्य
श्र	पारिपाश्वक
ष्क्र	निष्क्रिय
ष्क्व	निष्क्वाथ
ष्ट्य	उक्तिवैशिष्ट्य
ष्ट्र	अंतरराष्ट्रीय
ष्ट्य	ओष्ट्र्य

ष्य	वैतृष्य
ष्र	निष्प्रयोजन
स्क्र	स्क्रिप्ट
स्क्व	मस्क्वा
स्क्व	स्क्वाइन
स्ट्र	इंडस्ट्रियल
स्त्य	अगस्त्य
स्त्र	अधोवस्त्र
स्त्व	अंतस्त्वचा
स्थ्य	स्वास्थ्य
स्प्र	स्प्रिंग
स्प्ल	एस्प्लेनेड
द्व्र	सद्व्रत

6.3.3.3.CHCHCHC (Combination of four Consonants)

These combinations are rare and found only in the following case⁴⁵:

स्त्र्य स्त्र्याजीव

6.3.3.4 CHCHCHCHC (Combination of five Consonants)

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

⁴⁵ 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 ं is sorted along with क, since the first letter of that conjunct is क and on a similar principle ञ 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	ऐ ⁴⁶	ऐ	ओ	ऑ
औ	क	ख	ग	घ	ड	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	হ		

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.

ঁ	ং	ঃ	অ	আ	ই	ঈ	উ	ঊ	ঋ	এ	ঐ
ঐ	ও	ঔ	ঔ	ক	খ	গ	ঘ	ড	চ	ছ	জ
ঝ	জ	ট	ঠ	ঢ	ণ	ণ	ত	থ	দ	ধ	ন
প	ফ	ব	ভ	ম	য	ৰ	ল	ও	শ	ষ	স
হ	া	ি	ি	়	়	়	়	়	়	়	়
ঁ	ঁ	ঁ	ঁ	ঁ	ঁ	ঁ	ঁ	ঁ	ঁ	ঁ	ঁ

⁴⁶ ঐ and ওঁ are normally not listed in traditional dictionaries, but are included as digital dictionaries do include these characters.

In Tabular format:

ੁ	ੂ	ਾ	ਅ	ਆ	ੴ	ੴ	ਤ	ਊ	ੜ	ਏ	ੋ
ਐ	ਓ	ਾਂ	ਐ	ਕ	ਖ	ਗ	ਘ	ਙ	ਚ	ਛ	ਜ
ੱ	ਜ	ਟ	ਠ	ਡ	ਛ	ਣ	ਤ	ਥ	ਦ	ਧ	ਨ
ਪ	ਫ	ਬ	ਭ	ਮ	ਯ	ਰ	ਲ	ਵ	ਸ਼	਷	ਸ
ਹ	ਾ	ਿ	ੀ	ੁ	ੂ	ੂ	ੇ	ੁ	ੈ	ੋ	ੋ
ਐ	੍	੍									

Following is the example of sort order for consonant “ਕ”

ਕ	ਕੁ	ਕੁ	ਕਾ	ਕਿ	ਕੀ	ਕੁ	ਕੂ	ਕੂ	ਕੇ	ਕੁ
ਕੈ	ਕੋ	ਕੌ	ਕੌ	ਕ੍	ਕ੍					

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

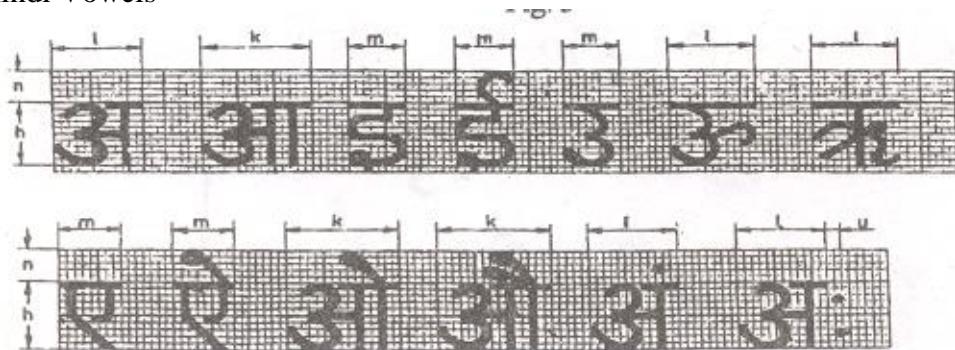
The first meeting of the “Expert Committee on Linguistic Resources and Language Technology Standards” for Validation of Hindi Script Grammar was held at DeitY on 12th Oct, 2012, under the chairmanship of Ms. Swaran Lata, HoD (HCC).

The list of participants :

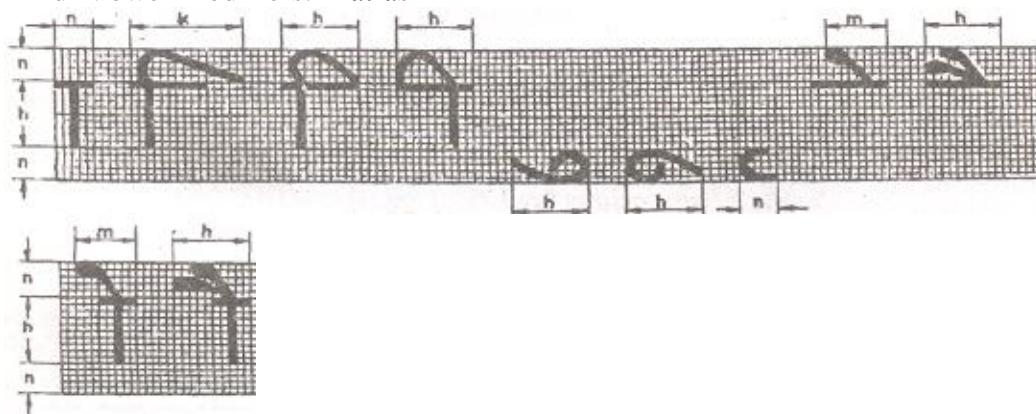
S. No.	Name	Organization	Mobile	Email
1.	Ms. Swaran Lata	DeitY	9868851170	słata@mit.gov.in
2.	Dr. Mukul Kumar Sinha	Expert Software consultants	9810874961	mukuls@yahoo.co.in
3.	Prof. Panchanan Mohanty	University of Hyderabad	9440034166	panchananmotianty@gmail.com
4.	Prof. B. Mallikarjun	Central University 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	tamuojha@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	vrij.nathan@gmail.com
32.	Prof. Pushpak Bhattacharya	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

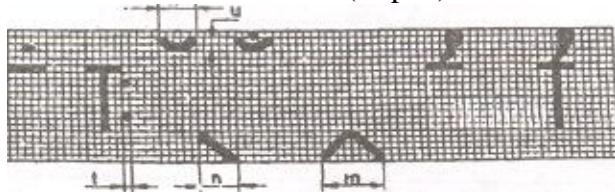
Annexure 2: Shapes of Hindi characters as per Central Hindi Directorate
Hindi Vowels



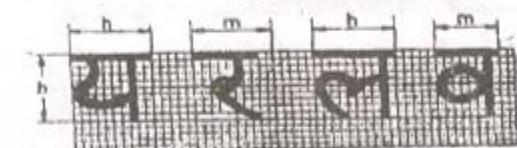
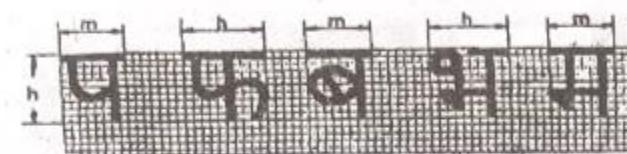
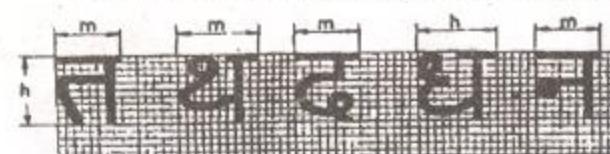
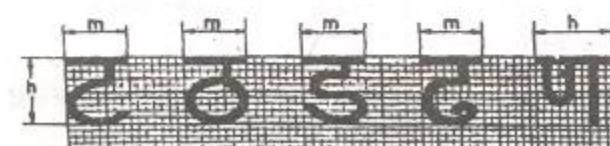
Hindi Vowel Modifiers: Mātrās



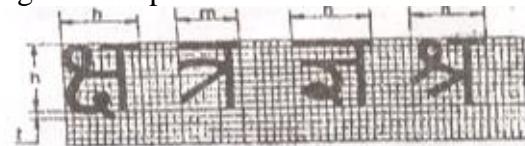
Hindi Diacritics and Rafar(Repha) Rakar



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:

शक्ति⁴⁷ 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.

⁴⁷ 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 पछा 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 Devanāgarī⁴⁸

⁴⁸ 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.
तं, स्तं, त्सं, त्स्यं, फङ्, वफङ्

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

त्, स्त्, फङ्

- zero or more consonant+(Nukta)+virāma sequences followed by a consonant (+Nukta) followed by a vowel sign is an akṣara, e.g.

र्ता, त्स्या, फङ्गी, लफङ्गी

- 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, स्त्रं, स्तः, फङ्

- 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.

त्स्या:, त्स्युं, त्स्युः, फङ्गे

- nothing else is an akṣara.

Annexure 6: NOTE ON AMBIGUOUS USE OF ANUSWARA (ANUNASIKA) WITH SUPERSCRIPT VOWEL SIGNS IN HINDI⁴⁹

This note deals with the ambiguity in the use of ‘bindu’ for both anuswāra and anunāsika with superscript vowel mātrās.

1. Notice that the ‘bindu’ stands for anunāsika on subscripted vowel mātrās: thus कं/ सं/ पुं can stand for kəm, kən, kəɳ, kəɳ and kəɳ/ sam, san, saɳ, saɳ and saɳ/ pum, pun, puɳ, puɳ and puɳ, respectively.

The ‘chandrabindu’ stands for anuswāra on subscripted vowel mātrās: thus कँ/ सँ/ पुँ stand for only kə/ sā/ pū, respectively.

However, for the superscripted vowel mātrās, the present-day Devnāgarī has only the ‘bindu’, which functions for both the anuswāra and the anunāsika: thus कों stands for both anuswāra (kō as in कोपल ‘bud’) and anunāsika (kom, kon, koɳ, koɳ, koɳ as in कोडा, कोंकण etc.). The examples can be given for all the four vowel mātrās represented with superscript diacritics, e.g सिंध ‘Sindh’ सिंचार्ड ‘irrigation’, etc.

2. In order for Devnāgarī to be used by other soliciting languages, it is necessary that this ambiguity is somehow recommended to be resolved. The superscript vowel mātrās should unambiguously go with the bindu for anunāsikas and chandrabindu for the anuswāra.

The proposed revision will be especially needed in the use of Devnāgarī for language technology for other languages in which both nasal vowels and nasal consonants occur as phonemes.

⁴⁹ Dr Pramod Pandey, Centre for Linguistics, J.N.U., New Delhi. (slightly modified version)

RECOMMENDATIONS FOR DEVELOPERS

The following guidelines are laid down for the developers. For the developer certain issues are preponderant. Since Indian scripts derived from Brahmi are syllable based, knowledge of the Indic Akshar is a must. As a corollary to the same, reduction of the Indic syllable to a basic formalism is a pre-requisite. This ensures not only correct cursor movement but when implemented also ensures that all incorrect syllables and syllable formation rules are flagged as such. Normalisation is also a major issue confronting the developer. On the other hand guidelines regarding norms as to correct usage, standardised spellings, rendering of loan words and punctuation are a must.

This document therefore addresses itself to these issues. It is divided into two major divisions.

In the first part, all issues afferent to the Akshar are handled. Part 1 which is technical in nature addresses the Indic syllable or the Akshar and provides a general overview of the same. This is followed by an overview of the formalism underlying the Akshar and finally issues derived from the Akshar such as correct and incorrect Akshar formation, cursor movement, normalisation. This part of the document is more a compendium since a majority of the issues raised here have been studied and addressed in different related forums viz. Unicode. Web references to these resources have been provided as footnotes or in the main body of the text.

In the second part, linguistic issues such as correct usage, spellings, loan words, numerals and punctuation are analysed. This part of the document reproduces in toto the guidelines as laid down by the Central Hindi Directorate in its *देवनागरी लिपि तथा हिंदी वर्तनी का मानकीकरण* New Delhi. 2010¹.

Each of these will be treated at length in what follows.

¹ Since a large part of this section is extracted verbatim from the guidelines laid down by the Central Hindi Directorate, these have been reproduced as such. It is requested that the requisite copyright permission be obtained from the Central Hindi Directorate for inclusion in this document

1. Technical issues

This part gives an overview of the Script, followed by a description of the Indic syllable: Akshar. These two sub-sections ensure that the developer is aware of the script as well as the basic building block of all Brahmi based scripts: Akshar. The next section provides a formalism for handling Akshar. This formalism is crucial since, by its application, valid and invalid syllables will be identified. It will also be pertinent to the sub-sections that treat of cursor movement and segmentation. Since Unicode in some cases has introduced more than one way of inputting a given conceptual character, Normalization of the text i.e. bringing the data in one uniform representation, is of prime importance.

1.1. DEVANAGARI SCRIPT AN OVERVIEW²

This overview is a linguistic introduction to Devanagari. It starts off with the historical evolution of Devanagari. In Section 1.3 it develops the notion of the underlying nucleus: the akshar and further draws attention to certain akshar structures relevant to variants. Simple transliteration has been used as a guide to the pronunciation of the examples.

1.2. Devanāgarī: A Historical Perspective

Devanagari is the main script for the Indo-Aryan languages Hindi, Marathi, Maithili, Dogri, Boro, Santhali, Sanskrit and Nepali recognized as official languages of the Republic of India. The script is also shared with other countries such as Fiji (Hindi) and Nepal (Nepali). It is the only script also for the related Indo-Aryan languages Bagheli, Bhili, Bhojpuri, Himachali dialects, Magahi, Newari and Rajasthani. It is associated closely with the ancient languages Sanskrit and Prakrit. It is an alternative script for Kashmiri (by Hindu speakers), Sindhi and Santhali. It is rising in use for speakers of tribal languages of Arunachal Pradesh, Bihar and Andaman & Nicobar Islands.

It is well-known that Devanagari has evolved from the parent script Brāhmī, with its earliest historical form known as Aśokan Brāhmī, traced to the 4th century B.C. Brāhmī was deciphered by Sir James Prinsep in 1837. The study of Brāhmī and its development has shown that it has given rise to most of the scripts in India, as mentioned above, and some outside India, namely, Sri Lanka, Myanmar, Kampuchea, Thailand, Laos, and Tibet.

The evolution of Brāhmī into present-day Devanagari involved intermediate forms, common to other scripts such as Gupta and Śāradā in the north and Grantha and Kadamba in the South. Devanagari can be said to have developed from the Kutila script, a descendant of the Gupta script, in turn a descendent of Brāhmī. The word kutila, meaning ‘crooked’, was used as a descriptive term to characterize the curving shapes of the script, compared to the straight lines of Brāhmī. A look at the development of Devanagari from Brāhmī gives an insight into how the Indic scripts have come to be diversified: the handiwork of engravers and writers who used different types of strokes leading to different regional styles.

²² Extracted from DEVANĀGARĪ VIP TEAM: VARIANT ISSUES REPORT pp. 7 et seq. This part of the document was drafted by Dr Pramod Pandey.
<https://community.icann.org/download/attachments/17466438/Devanagari+VIP+Team+Issues+Report+1.0.docx?version=1&modificationDate=1315604624000>.

1.3 Encoding Principles

Chapter 9 of Unicode 6.1³ in its Devanagari section clearly lays down the encoding principles underlying Devanagari script. The relevant pages 211-215 give an overview of the basic structure of Devanagari and cover the following areas:

- Consonant Letters.
- Independent Vowel Letters
- Dependent Vowel Signs (Matras).
- Virama, Halant, or Vowel Omission Sign
- Consonant Conjuncts

The properties underlying each of these are treated at length in this document.

<http://www.unicode.org/versions/Unicode6.1.0/ch09.pdf>. Owing to copyright reasons, the text in the document cannot be incorporated in this document. Developers can access the document for details. This applies to all Unicode documents which are protected by copyright laws.

1.4 The Fundamental Unit: akshar⁴

The classification of Devanagari characters as seen in the document can be reduced to a “compositional grammar” based on a Backus-Naur formalism (ISCII ’91 IS 13194) which ensures the well-formedness of the akshar. The formalism describes the nodal units of the script: Consonant and Vowel and determines which elements can be conjoined to each of these Nodal Units. The fundamental properties of the akshar are defined below:

The akshar is the graphemic unit of Devanagari. The difference between the syllable and the akshar is that while the syllable includes one or more post-vocalic consonants, the akshar doesn’t, as can be seen below:

Phonemic forms	Syllabic units	Akshar units
cha:ruləta:	CV.CV.CV.CV	CV. CV. CV. CV
e:k	VC.	V. C
upka:r	VC. CVC	V. C. CV. C
indira	VC. CV. CV	VC. CV. CV
əst	VCC	V. CC
əkʃər	VC. CVC	V. CCV. C

Table 1: Syllabic and akshara divisions of spoken forms

As can be seen from Table 1, there is a marked difference between the written and spoken syllable, especially insofar as the division of consonant clusters across syllable boundaries e.g. /upka:r/ is concerned.

The only exception to the generalization about the post-vocalic consonants vis-à-vis akshar is the anuswāra, the underlying nasal consonant surfacing as homorganic with the following stop. The anuswāra is treated as a part of the grapheme. The orthographic and phonetic transcriptions of forms with the anuswāra are given below:

बिंदी	[bindi:]	‘point _N ’
कंबल	[kəmbəl]	‘blanket _N ’
डंडा	[dəɳɖa:]	‘stick _N ’
खंजर	[k ^h əɳɟər]	‘knife _N ’
कंघी	[kəɳg ^h i:]	‘comb _N ’

Table 2: Representation of anuswāra in Devanāgarī

The vowel is an independent unit of akshar word-initially and post-vocallyically.

अ	आ	इ	ई	उ	ऊ	ए	ऐ	ओ	औ
ə	a	i	i:	u	u:	e:	æ:	o:	əu

Table 3: Independent vowel letters

⁴ Extracted from DEVANĀGARĪ VIP TEAM: VARIANT ISSUES REPORT pp. 7 et seq. This part of the document was drafted by Dr Pramod Pandey

Vowels and consonants are assumed to be different types of units and are so represented in the grapheme when the vowels follow consonants. The following akshar consist of single consonants followed by a vowel:

क	का	कि	की	कु	कू	के	कै	को	कौ
kə	ka	ku	ki:	ku	ku:	ke:	kæ:	ko:	kəu

Table 4: *Devanāgarī CV akshar*

As can be seen in the first grapheme in Table 3, the neutral vowel /ə/ is assumed to be inherent in a consonant. The vowel is pronounced as such word initially and medially in certain contexts, for example, in the first grapheme in पल /pəl/. The inherent neutral vowel is not pronounced word-finally or medially in certain contexts.

Two-consonant clusters

The following features can be attributed to the two consonant clusters:

1. Generally, half the letter of the first consonant precedes the full letter of the second consonant: e.g., स्क<sk>, प्त<pt>, क्ल<kl> etc. Alternatively, the practice of specifying the diacritic for unreleased consonants, known as ‘halanta’, is used for the first consonant, e.g., द्भ<db^h>उद्भव्य/u db^həv/
2. For a C+r cluster, as noted above, the /r/ is specified as a subscript that looks like an inscript: क्र<kr>, ख्र<k^hr>, फ्र<p^hr>.
3. For r+C clusters, the /r/ is specified as a superscript above the grapheme, e.g., र्म<rm>, र्ज<rdʒ>
4. In the case of the following two-consonant clusters, a new ligatured group is formed. These are: त्र<tr>, ष्ट<kʃ>, ज्ञ<jŋ>, श<jr>, क्त<kt>.

Three-consonant clusters:

These admit the following features:

1. Generally, the first two consonants are specified for half their letters, and the third is fully specified, e.g., स्प्ल<spl>. This convention is usually followed for borrowed words.
2. For C+C+r clusters, and for r+C+C clusters, which are highly restricted, the convention for two-consonant clusters applies, e.g., स्त्र<str>

Four-consonant clusters:

These are by far rare and have been listed in the document : *Devanagari Script Behaviour for Hindi..*

1.5 FORMALISM: DEFINITION OF THE INDIC AKSHAR⁵

The formalism governing the Akshar as pertinent to Hindi is provided below. The formalism was provided from the W3C site. It explains the underlying complexities of the Indic Akshar formalism. The formalism based on a modified Backus-Naur Formalism accommodates all valid akshars and by very definition demarcates all syllables which are invalid. Correct implementation of the rules laid down in the document would ensure that all valid akshar in Hindi would be defined.

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, इ, 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.
त्, त्स्, त्व्, त्व्य्, फ्क्, वफ्क्
- **zero or more consonant(+Nukta)+virāma sequences followed by a consonant (+Nukta) followed by a virāma is an akṣara**, e.g.
त्ू, स्त्ू, फ्क्ू
- **zero or more consonant+(Nukta)+virāma sequences followed by a consonant (+Nukta) followed by a vowel sign is an akṣara**, e.g.
रा, त्व्या, फ़ी, लफ़ी
- **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.

⁵ http://www.w3cindia.in/Indic_Aksara.html#need Since this is an excerpt from the document the term akshar written as **akṣara** is retained as such although Unicode prescribes akshar

स्तं, स्त्रं, स्तः, फ़्ऱं

- 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.
त्स्न्या:, त्स्न्युं, त्स्न्यु॒, फ़्ऱे॑
- nothing else is an akṣara.

1.6. VALID AND INVALID AKSHAR FORMATION:

The notion of valid and invalid Akshar formation necessarily devolves from the Akshar formation rules. A couple of examples will illustrate the formalism at work. Let us take the case of Vowels.

Rules

Two rules define a valid Vowel akshar formalism

- **any independent vowel is an akṣara**, e.g.

आ, ई, उ

- **any independent vowel followed by anusvāra, candrabindu or visarga is an akṣara**, e.g. अँ, उँ, आः

A closing rule

- **nothing else is an akṣara**

clearly states that all other combinations are invalid syllables.

Explanation of the Rules

Rule 1 states that a vowel by itself is an akshar. This rule is supplemented by Rule 2 which states that an independent vowel followed by anuswara, candrabindu or visarga constitutes a valid akshar. Rule 3 states that nothing else is a “vowel” akshar, thereby delimiting vowel akshar to the following cases only:

आ

आँ

अँ

आः

Application

By definition, therefore, in strings such as :

अँूं

आःूं

उँूं

invalid combinations are present and this because

अँूं Candrabindu and anuswar cannot be present sequentially

आःूं Visarga cannot be followed by a candrabindu

उँूं Anuswar cannot be followed by a candrabindu

Similar rules applicable for consonants determine what constitutes a valid consonant akshar and precludes cases such as

काौ - A consonant followed by a Vowel sign cannot be followed by a second vowel sign

गौौ - A consonant followed by a candrabindu cannot be followed by a vowel sign

फःौ - A consonant followed by Visarga cannot be followed by anuswara

1.7. UNICODE TEXT SEGMENTATION UAX #29⁶

The notion of akshar is essential to this topic which define “*guidelines for determining default boundaries between certain significant text elements: user-perceived characters, words, and sentences*⁷. ”

This process whereby boundaries are delimited results in segmentation. Unicode Standard Annexure UAX#29 which has undergone many changes provides clear-cut guidelines for text segmentation.

Since the document is self sufficient and defines very clearly all and every case of segmentation, it will not be treated at length. However a few examples of the manner in which akshar allows for syllable segmentation will illustrate the importance of segmentation.

Given a string such as:

मनमोहन ओष्ठ्य

The akshar rules allow for clear intra-word segmentation and delimit each word into its valid syllables

म - न - मो - ह - न

C - C - CM -C - C

ओ - ष्ठ्य

V - C+H+C+H+C

This division clearly shows that the first word can be broken up into 5 syllables whereas the second has just 2 syllables.

As an extension of the above, a word comprising one or more akshar followed by space followed a number or another string comprising one or more akshar shall constitute two words

राम आया

UAX#29 provides in detail all such rules pertinent to text segmentation.

⁶ <http://www.unicode.org/reports/tr29/>

⁷ Op cit.

1.8 RENDERING

Closely allied to the issues studied above is the problem of rendering. Chapter 9 of Unicode, South Asian Scripts - I studies in depth the issue of rendering in Devanagari.

“Devanagari characters, like characters from many other scripts, can combine or change shape depending on their context. A character's appearance is affected by its ordering with respect to other characters, the font used to render the character, and the application or system environment. These variables can cause the appearance of Devanagari characters to differ from their nominal glyphs (used in the code charts).

Additionally, a few Devanagari characters cause a change in the order of the displayed characters. This reordering is not commonly seen in non-Indic scripts and occurs independently of any bidirectional character reordering that might be required.”⁸

The Chapter⁹ focuses on the following issues pertinent to rendering:

- Rendering Consonant Letters.
- Rendering Independent Vowel Letters
- Rendering Dependent Vowel Signs (Matras).
- Rendering Virama
- Rendering Consonant clusters which includes the following:
 - conjunct formation
 - dead consonants
- Explicit Virama for preventing conjunct forms, generating explicit half-consonants
 - and most importantly
- Rendering rules

Although the use of ZWJ/ZWNJ is pre-implied in the chapter, it will be treated separately in 1.9

⁸ <http://www.unicode.org/versions/Unicode6.1.0/ch09.pdf>. p. 212
⁹ pp. 211-22

1.9. Zero Width Joiner(ZWJ)/Zero Width Non-Joiner(ZWNJ) IN HINDI

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.

However as a case in point, Unicode 2.0 prescribes the use of RA+VIRAMA+ZWJ to represent the eyelash-ra used in Marathi and Konkani and Nepali. 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 *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. Thus the use of ZWJ/ZWNJ is restricted to the display level.

RULE 1: If a consonant+halant is followed by the ZWJ, the half-form of the consonant is formed.

Example:

शक्ति¹⁰ 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.

¹⁰ As in the case of Marathi

RULE 2: The use of ZWNJ in Hindi is restricted to representing a dead consonant within a string.

Thus to show the combination of राजू+कमल as a single word and retain the shape of the consonant followed by the halant; ZWNJ is used:

राजूकमल : 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

RULE 3: 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 from the above, 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, since storage is affected, except in the case of Sanskrit loans where such usage is mandatory, as in Rule 2 above.

1.10. CURSOR MOVEMENT & DELETION

1.10.1 Cursor movement

Cursor movement is an important issue in content creation. Given the complex nature of Brahmi based scripts as opposed to Latin; the developer should be aware of cursor movement which affects not only word-processing but underlies all display.

Cursor movement in the case of Latin or Cyrillic scripts moves the cursor across one character at a time. Thus in a string such as

Enter

The cursor moving either from right to left or left to right will move across one character at a time.

If Ctrl+Right Arrow or Ctrl+Left arrow are pressed, the cursor will move across the word. Pressing Home or End will move the cursor to the beginning or end of the line respectively.

Ctrl+Home or CTRL+End will move the cursor to the beginning of the document or the end of the document respectively.

In the case of Brahmi based scripts and more specifically Devanagari in this instance, the cursor movement is slightly different as the following quote¹¹ explains:

For Latin script, there is often a direct one-to-one mapping between a character and its glyph. (For instance, the character "h" is always represented by the same glyph "h.") For complex scripts, several characters can combine together to create a whole new glyph independent of the original characters. There are also cases where the number of resulting glyphs can be bigger than the original number of characters used to generate those glyphs. Characters are often stacked or combined to create a cluster, which is indivisible for most of the complex scripts.

In the case of Devanagari, the cursor will move across syllables and not across the individual characters.

The Table below shows the basic difference between Devanagari and Latin

MOVEMENT	LATIN	DEVANAGARI
Right arrow	To right of character	To right of syllable
Left Arrow	To left of character	To left of syllable
Ctrl+Right arrow/Home key	Following word	Following word

¹¹ <http://msdn.microsoft.com/en-us/goglobal/bb688137.aspx>

Ctrl+Left arrow	Preceding word	Preceding word
Home Key	Beginning of line	Beginning of line
End Key	End of Line	End of Line
Ctrl+Home	Beginning of Document	Beginning of Document
Ctrl+End	End of Document	End of Document

A special case is that of the insertion of ZWJ/ZWNJ within the syllable. In that case the behaviour is as under

Rule 1: When a ZWJ is inserted within a syllable the cursor, the cursor treats that particular syllable as a single unit and moves across it in either direction as in the case of

रक्त U+0930 U+0915 U+094D U+200D U+0924

Rule 2: When a ZWJ is inserted within a syllable the cursor, the cursor treats the particular syllable slightly differently. The syllable is broken at the point where the ZWNJ is inserted and is virtually treated as two syllables, as in the case of

रकृत U+0930 U+0915 U+094D U+200C U+0924

1.10.2 DELETION

In the case of deletion, a similar difference exists between Latin and Brahmi based scripts. In the case of Latin scripts, deletion is across characters. Thus pressing BACKSPACE after positioning the cursor before the character/space to be deleted, deletes the character/space to the left. Pressing the DEL key on the other hand will delete the character/space to the right.

In the case of Devanagari which has the Akshar as the basic unit, the situation is different.

Pressing BACKSPACE after positioning the cursor before the character/space results in each successive character of the syllable being deleted.

रक्त रक् रक र

Pressing the DEL key produces a different effect, this time the full syllable is deleted:

रक्त क्त

Pressing Ctrl+DEL or CTRL+BACKSPACE deletes the full word.

In the case of ZWJ inserted within the syllable, the following has to be noted
 Rule: Given a syllable with a ZWJ inserted, BACKSPACE will delete a single character at a time, as shown below.

रक्त रक् रक् रक र

Rule: Given a syllable with a ZWJ inserted, DEL will delete the full syllable:

रक्त क्त

In the case of ZWNJ inserted within the syllable, the following has to be noted

Rule: Given a syllable with a ZWNJ inserted, BACKSPACE will delete a single character at a time, as shown below.

रक्ृत रक् रक् रक् र

Rule: Given a syllable with a ZWNJ inserted, DEL will treat the halant+ZWNJ as a single character and delete it as shown below

रक्ृत कृत त

1.11 NORMALIZATION

A major issue in Hindi is that of Normalization. Normalization is of such importance that the document Unicode Standard Annex #15¹² has undergone multiple changes.

In simple terms, Normalization has to take place when a given conceptual character can be rendered in two or more ways. Since such multiple representation can affect searching, sorting (and lead to spoofing and phishing), it is essential that these multiple representations be reduced to one single canonical form.

Unicode provides useful charts for the most common Normalization forms in all scripts¹³. In the case of Hindi the Normalization is seen mainly in the case of the Nukta characters. Unicode 3.0 upwards permits two different ways of representing the 6 Nukta characters of Hindi: akhand form vs. Consonant+Nukta.

Thus

ऋ U+0958 ॠ U+0915 U+093C
ঢ় U+095C ঢ় U+0921 093C

The akhand form is preferred to the combination Consonant+nukta as is shown in the table below¹²:

Normalized form	Combination to be normalized
ঢ় 0958	ঢ় 0915 093C
ঢ় 0959	ঢ় 0916 093C
ঢ় 095A	ঢ় 0917 093C
ঢ় 095B	ঢ় 091C 093C
ঢ় 095C	ঢ় 0921 093C
ঢ় 095D	ঢ় 0922 093C

Table 5

¹² <http://unicode.org/reports/tr15/>

¹³ <http://www.unicode.org/charts/normalization/>

In the case of web-browsers within the address bar, a URL containing a word with Consonant+nukta is automatically normalised to its corresponding normalized(Akhand) form due to punycode algorithm that gets into act while accessing Internationalized Domain Names. Thus:

अङ्ग.com U+092C U+0922 U+093C.com
gets automatically normalized to
अङ्ग.com U+092C U+095D

Developers would do well to study the document to understand the complexity of the problem along with the FAQ on Normalization¹⁴ as well as the Normalization, Sorting and Searching sections of chapter 5 of the Unicode Standard¹⁵. Detection and Testing of Normalization forms is also provided in useful links¹⁶. International Components for Unicode (ICU) also supports related, additional operations. Some of them are described in Unicode Technical Note #5 (Canonical Equivalence in Applications)¹⁷.

¹⁴ <http://www.unicode.org/faq/normalization.html>. This document though not pertinent to Hindi provides an excellent overview of Normalization. It provides Normalization Charts for Scripts.

¹⁵ <http://www.unicode.org/versions/Unicode6.2.0/>

¹⁶ Normalization forms detection

http://unicode.org/reports/tr15/#Detecting_Normalization_Forms,

Normalization test

<http://www.unicode.org/Public/UNIDATA/NormalizationTest.txt>

¹⁷ <http://www.unicode.org/notes/tn5/>

2. LINGUISTIC ISSUES

2.1 PRELIMNARIES

This section is based on the document “देवनागरी लिपि तथा हिंदी वर्तनी का मानकीकरण – 2010” of the Central Hindi Directorate pp. 20 - 55.

The guidelines laid down therein cover the following areas of interest to developers. These can be divided into the following parts. Since fidelity to the original text is of prime importance and should not be *lost in translation*, the recommendations have been reproduced verbatim.

These guidelines and recommendations constitute a style guide which covers wide range of areas: correct usage, norms for spellings of loan words, punctuation, correct spellings of numbers and fractions and also paragraph formatting.

2.2. CORRECT USAGE

a. Recommendations for affixing the कारक चिह्न 3.2. of the document.

Case marker adjunction is important in Hindi since case markers can be adjuncted to pronouns but are written separate from the Noun. However if the pronoun is followed by ही तक, the case marker is written separate as in आप ही के लिए। This section also stipulates that colloquial forms such as मेरेको, मेरेसे be avoided.

हिंदी के कारक चिह्न सभी प्रकार के संज्ञा शब्दों में प्रतिपदिक से पृथक् लिखे जाएँ। जैसे :- राम ने, राम को, राम से, स्त्री का, स्त्री से, सेवा में आदि। सर्वनाम शब्दों में ये चिह्न प्रतिपदिक के साथ मिलाकर लिखे जाएँ। जैसे :- तूने, आपने, तुमसे, उसने, उसको, उससे, उसपर, मुझको, मुझसे आदि (मेरेको, मेरेसे आदि रूप व्याकरण सम्मत नहीं हैं)।

सर्वनामों के साथ यदि दो कारक चिह्न हों तो उनमें से पहला मिलाकर और दूसरा पृथक् लिखा जाए। जैसे :- उसके लिए, इसमें से।

सर्वनाम और कारक चिह्न के बीच 'ही', 'तक' आदि निपात हों तो कारक चिह्न को पृथक् लिखा जाए। जैसे :- आप ही के लिए, मुझ तक को।

b. Recommendations for Hyphenation Section 3.4. of the document.

Hyphenation is an area little touched upon in Indian languages. The Central Hindi Directorate lays down precise guide-lines for hyphenation, which are reproduced below;

योजक चिह्न (हाइफन) का विधान स्पष्टता के लिए किया गया है।

द्वंद्व समास में पदों के बीच हाइफन रखा जाए। जैसे :- राम-लक्ष्मण, शिव-पार्वती संवाद, चाल-चलन, हँसी-मज़ाक, लेन-देन, पढ़ना-लिखना, खाना-पीना, खेलना-कूदना आदि।

सा, से, सी आदि से पूर्व हाइफन रखा जाए जैसे :- तुम-सा, चाकू-से तीखे।

तत्पुरुष समास में हाइफन का प्रयोग केवल वहीं किया जाए जहाँ उसके बिना भ्रम होने की संभावना हो, अन्यथा नहीं। जैसे :- भू-तत्व। सामान्यतः तत्पुरुष समास में हाइफन लगाने की आवश्यकता नहीं है। जैसे :- रामराज्य, राजकुमार, गंगाजल, ग्रामवासी, आत्महत्या आदि।

इसी तरह यदि 'अ-नख' (बिना नख का) समस्त पद में हाइफन न लगाया जाए तो उसे 'अनख' पढ़े जाने से 'क्रोध' का अर्थ भी निकल सकता है। अ-नति (नम्रता का अभाव) : अनति (थोड़ा), अ-परस (जिसे किसी ने न छुआ हो) : अपरस (एक चर्म रोग), भू-तत्व (पृथ्वी-तत्व) : भूतत्व (भूत होने का भाव) आदि समस्त पदों की भी यही स्थिति है। ये सभी युग्म वर्तनी और अर्थ दोनों दृष्टियों से भिन्न-भिन्न शब्द हैं।

कठिन संधियों से बचने के लिए भी हाइफन का प्रयोग किया जा सकता है। जैसे :- दृवि-अक्षर न कि दव्यक्षर; दृवि-अर्थक न कि दव्यर्थक, त्रि-अक्षर न कि त्र्यक्षर आदि।

c. Recommendations for अनुस्वार and चंद्रबिंदु: Section 3.6 of the document.

Both Anuswar and Chandrabindu shall be used in Hindi. However the usage of these two shall be governed by precise guide-lines which are stipulated at length and are reproduced below.

अनुस्वार व्यंजन है और अनुनासिक स्वर का नासिक्य विकार। हिंदी में ये दोनों अर्थभेदक भी हैं। अतः हिंदी में अनुस्वार (—) और अनुनासिकता चिह्न (—) दोनों ही प्रचलित रहेंगे।

Guidelines concerning Anuswar.

The following guidelines govern the use of Anuswar:

अनुस्वार (शिरोबिंदु —)

संस्कृत शब्दों में अनुस्वार का प्रयोग अन्यवर्गीय वर्णों ('य' से 'ह' तक) से पहले यथावत् रहेगा। जैसे :- संयोग, संरक्षण, संलग्न, संवाद, अंश, कंस, सिंह आदि।

संयुक्त व्यंजन के रूप में जहाँ पंचम वर्ण (पंचमाक्षर) के बाद सर्वर्गीय शेष चार वर्णों में से कोई वर्ण हो तो एकरूपता और मुद्रण/लेखन की सुविधा के लिए अनुस्वार का ही प्रयोग करना चाहिए। जैसे :- पंकज, गंगा, चंचल, कंजूस, कंठ, ठंडा, संत, संध्या, मंदिर, संपादक, संबंध आदि (पङ्कज-पङ्कज, गङ्गा-गङ्गा, चञ्चल, कञ्जूस, कण्ठ, ठण्डा, सन्त, मन्दिर, सन्ध्या, सम्पादक, सम्बन्ध वाले रूप नहीं)।

यदि पंचमाक्षर के बाद किसी अन्य वर्ग का कोई वर्ण आए तो पंचमाक्षर अनुस्वार के रूप में परिवर्तित नहीं होगा। जैसे :- वाड्मय, अन्य, चिन्मय, उन्मुख आदि (वांमय, अंय, चिंमय, उंमुख आदि रूप ग्राह्य नहीं होंगे)।

पंचम वर्ण यदि द्वितीय रूप में (साथ-साथ) आए तो पंचम वर्ण अनुस्वार में परिवर्तित नहीं होगा। जैसे :- अन्, सम्मेलन, सम्मति आदि (अंन, संमेलन, संमति रूप ग्राह्य नहीं होंगे)।

अंग्रेजी, उर्दू से गृहीत शब्दों में आधे वर्ण या अनुस्वार के भ्रम को दूर करने के लिए नासिक्य व्यंजन को पूरा लिखना अच्छा रहेगा। जैसे :- लिम्का, तन्खाह, तिन्का, तम्गा, कम्सिन आदि।

संस्कृत के कुछ तत्सम शब्दों के अंत में अनुस्वार का प्रयोग म् का सूचक है। जैसे :- अहं (अहम्), एवं (एवम्), परं (परम्), शिवं (शिवम्)।

Guidelines for use of Chandrabindu.

The following are the guidelines governing the use of Chandrabindu:

अनुनासिक चिह्न (चंद्रबिंदु ॒)

हिंदी के शब्दों में उचित ढंग से चंद्रबिंदु का प्रयोग अनिवार्य होगा।

अनुनासिक चिह्न व्यंजन नहीं है, स्वरों का ध्वनिगुण है। अनुनासिक स्वरों के उच्चारण में मुँह और नाक से हवा निकलती है। जैसे :— आँ, ऊँ, ऐँ, माँ, हूँ, आएँ।

चंद्रबिंदु के प्रयोग के बिना प्रायः अर्थ में भ्रम की गुंजाइश रहती है। जैसे :— हंस : हँस, अंगना : अँगना, स्वांग (स्व+अंग) : स्वाँग आदि में। अतएव ऐसे भ्रम को दूर करने के लिए चंद्रबिंदु का प्रयोग अवश्य किया जाना चाहिए। किंतु जहाँ (विशेषकर शिरोरेखा के ऊपर जुड़ने वाली मात्रा के साथ) चंद्रबिंदु के प्रयोग से छपाई आदि में बहुत कठिनाई हो और चंद्रबिंदु के स्थान पर बिंदु का (अनुस्वार चिह्न का) प्रयोग किसी प्रकार का भ्रम उत्पन्न न करे, वहाँ चंद्रबिंदु के स्थान पर बिंदु के प्रयोग की छूट रहेगी। जैसे :— नहीं, में, मैं आदि। कविता आदि के प्रसंग में छंद की दृष्टि से चंद्रबिंदु का यथास्थान अवश्य प्रयोग किया जाए। इसी प्रकार छोटे बच्चों की प्रवेशिकाओं (पुस्तकों) में जहाँ चंद्रबिंदु का उच्चारण अभीष्ट हो, वहाँ मोटे अक्षरों में उसका यथास्थान सर्वत्र प्रयोग किया जाए। जैसे :— कहाँ, हँसना आँगन, सँवारना, मैं, मैं, नहीं आदि।

d. Recommendations regarding the Visarga: Section 3.7 of the document.

The use of Visarga is an area prone to error, not only because it is mistaken for colon and written as such, but also because the intricacies of the use of the Visarga are little known. Below are given the major guidelines laid down by the Central Hindi Directorate and which describe where the Visarga sign should be used.

Very often in the case of निः and दुः the visarga is replaced by a doubling of the consonant as in दुःसाहस and निःशब्द which are acceptable. However, writing

निःस्वार्थ as निस्स्वार्थ is not acceptable. Below are provided guidelines as to the correct usage of the Visarga.

संस्कृत के जिन शब्दों में विसर्ग का प्रयोग होता है, वे शब्द यदि तत्सम रूप में प्रयुक्त हों तो विसर्ग का प्रयोग अवश्य किया जाए। जैसे :- 'दुःखानुभूति' में। यदि उस शब्द के तद्भव रूप में विसर्ग का लोप हो चुका हो तो उस रूप में विसर्ग के बिना भी काम चल जाएगा। जैसे :- 'दुख-सुख के साथी'।

तत्सम शब्दों के अंत में प्रयुक्त विसर्ग का प्रयोग अनिवार्य है। जैसे :- अतः, पुनः, स्वतः, प्रायः, पूर्णतः, मूलतः, अंततः, वस्तुतः, क्रमशः आदि।

'ह' का अघोष उच्चरित रूप विसर्ग है, अतः उसके स्थान पर (स)घोष 'ह' का लेखन किसी हालत में न किया जाए (अतः, पुनः आदि के स्थान पर अतह, पुनह आदि लिखना अशुद्ध वर्तनी का उदाहरण माना जाएगा)।

दुःसाहस/दुस्साहस, निःशब्द/निशशब्द के उभय रूप मान्य होंगे। इनमें द्वित्व वाले रूप को ग्राथमिकता दी जाए।

निःस्वार्थ मान्य है। (निस्स्वार्थ उचित नहीं होगा)।

निस्तेज, निर्वचन, निश्चल आदि शब्दों में विसर्ग वाला रूप (निःतेज, निःवचन, निःचल) न लिखा जाए।

अंतःकरण, अंतःपुर, दुःस्वप्न, निःसंतान, प्रातः काल आदि शब्द विसर्ग के साथ ही लिखे जाएँ।

तद्‌भव/देशी शब्दों में विसर्ग का प्रयोग न किया जाए। इस आधार पर छः लिखना गलत होगा। छह लिखना ही ठीक होगा।

प्रायद्वीप, समाप्तप्राय आदि शब्दों में तत्सम रूप में भी विसर्ग नहीं है।

विसर्ग को वर्ण के साथ मिलाकर लिखा जाए, जबकि कोलन चिह्न (उपविराम :) शब्द से कुछ दूरी पर हो। जैसे :- अतः , यों है :-

e. Recommendations regarding use of the Halant: Section 3.8. of the document.

The Central Hindi Directorate states that the term Hal Cihna be used instead of Halant. Apart from the use of an explicit representation of the Halant for forming conjuncts where the first letter is ड,छ,ट,ठ,द,হ, the guidelines also amplify and provide recommendations for the use of the Halant.

() को हल् चिह्न कहा जाए न कि हलंत। व्यंजन के नीचे लगा हल् चिह्न उस व्यंजन के स्वर रहित होने की सूचना देता है, यानी वह व्यंजन विशुद्ध रूप से व्यंजन है। इस तरह से 'जगत्' हलंत शब्द कहा जाएगा क्योंकि यह शब्द व्यंजनांत है, स्वरांत नहीं।

संयुक्ताक्षर बनाने के नियम 3.1.2.2 के अनुसार ड् छ् ट् ट् द् द् ह् में हल् चिह्न का ही प्रयोग होगा। जैसे :- चिह्न, बुड़ा, विद्वान् आदि में।

तत्सम शब्दों का प्रयोग वांछनीय हो तो हलंत रूपों का ही प्रयोग किया जाए; विशेष रूप से तब जब उनसे समस्त पद या व्युत्पन्न शब्द बनते हों। जैसे :- प्राक्-(प्रागैतिहास), वाक्-(वाग्देवी), सत्-(सत्साहित्य), भगवन्-(भगवद्भक्ति), साक्षात्-(साक्षात्कार), जगत्-(जगन्नाथ), तेजस्-(तेजस्वी), विद्युत्-(विद्युल्लता) आदि। तत्सम संबोधन में हे राजन्, हे भगवन् रूप ही स्वीकृत होंगे। हिंदी शैली में हे राजा, हे भगवान लिखे जाएँ। जिन शब्दों में हल् चिह्न लुप्त हो चुका हो, उनमें उसे फिर से लगाने का प्रयत्न न किया जाए। जैसे :- महान्, विद्वान् आदि; क्योंकि हिंदी में अब 'महान्' से 'महानता' और 'विद्वान्' से 'विद्वानों' जैसे रूप प्रचलित हो चुके हैं।

व्याकरण ग्रंथों में व्यंजन संधि समझाते हुए केवल उतने ही शब्द दिए जाएँ, जो शब्द रचना को समझने के लिए आवश्यक हों (उत् + नयन = उन्नयन, उत् + लास = उल्लास) या अर्थ की दृष्टि से उपयोगी हों (जगदीश, जगन्माता, जगज्जननी)।

हिंदी में हृदयंगम (हृदयम् + गम), उद्धरण (उत्/उद् + हरण), संचित (सम् + चित्) आदि शब्दों का संधि-विच्छेद समझाने की आवश्यकता प्रतीत नहीं होती। इसी तरह 'साक्षात्कार' 'जगदीश' 'षट्कोण' जैसे शब्दों के अर्थ को समझाने की आवश्यकता हो तभी उनकी संधि का हवाला दिया जाए। हिंदी में इन्हें स्वतंत्र शब्दों के रूप में ग्रहण करना ही अच्छा होगा।

f. Recommendations for use of ऐ & औ:Section 3.10 of the document.

The use of the diphthongs ऐ & औ and the glides य and व are often a major source of error. The Central Hindi Directorate clearly specifies that forms such as कौव्वा गवच्चा i.e. geminating the glides be avoided and the words be written without such germination: कौवा गवया

हिंदी में ऐ (ऐ), औ (औ) का प्रयोग दो प्रकार के उच्चारण को व्यक्त करने के लिए होता है। पहले प्रकार का उच्चारण 'है', 'और' आदि में मूल स्वरों की तरह होने लगा है; जबकि दूसरे प्रकार का उच्चारण 'गवया', 'कौवा' आदि शब्दों में संध्यक्षरों के रूप में आज भी सुरक्षित है। दोनों ही प्रकार के उच्चारणों को व्यक्त करने के लिए इन्हीं चिह्नों (ऐ, ऐ; औ, औ) का प्रयोग किया जाए। 'गवच्चा', 'कव्वा' आदि संशोधनों की आवश्यकता नहीं है। अन्य उदाहरण हैं - भैया, सैयद, तैयार, हौवा आदि।

दक्षिण के अव्यर, नव्यर, रामच्चा आदि व्यक्तिनामों को हिंदी उच्चारण के अनुसार ऐयर, नैयर, रामैया आदि न लिखा जाए, क्योंकि मूलभाषा में इसका उच्चारण भिन्न है।

अच्छल, कच्चाल, कच्चाली जैसे शब्द प्रचलित हैं। इन्हें लेखन में यथावत् रखा जाए।

संस्कृत के तत्सम शब्द 'शच्या' को 'शैया' न लिखा जाए।

g. Miscellaneous Recommendations

In its effort to ensure correct usage, the Central Hindi Directorate also lays down rules and guidelines for correct usage of कर वाला and when these should be joined to the word and when they should be separated from the word. Thus in the case of वाला, the guideline is that वाला shall be written separate when it is adjuncted to the verb to show intentionality, but shall be joined to the noun when it functions as an agentive suffix. The guidelines also distinguish between delicate cases such as

गाँववाला villager vs. गाँव वाला मकान village house

i. कर

पूर्वकालिक कृदंत प्रत्यय 'कर' क्रिया से मिलाकर लिखा जाए। जैसे :- मिलाकर, खा-पीकर, रो-रोकर आदि।

कर + कर से 'करके' और करा + कर से 'कराके' बनेगा।

ii. वाला

क्रिया रूपों में 'करने वाला', 'आने वाला', 'बोलने वाला' आदि को अलग लिखा जाए। जैसे :- मैं घर जाने वाला हूँ, जाने वाले लोग।

योजक प्रत्यय के रूप में 'घरवाला', 'टोपीवाला' (टोपी बेचने वाला), दिलवाला, दूधवाला आदि एक शब्द के समान ही लिखे जाएँगे।

'वाला' जब प्रत्यय के रूप में आएगा तब तो 3.12.2 के अनुसार मिलाकर लिखा जाएगा; अन्यथा अलग से। यह वाला, यह वाली, पहले वाला, अच्छा वाला, लाल वाला, कल वाली बात आदि में 'वाला' निर्देशक शब्द है। अतः इसे अलग ही लिखा जाए। इसी तरह लंबे बालों वाली लड़की, दाढ़ी वाला आदमी आदि शब्दों में भी वाला अलग लिखा जाएगा। इससे हम रचना के स्तर पर अंतर कर सकते हैं। जैसे :-

गाँववाला - villager गाँव वाला मकान - village house

b. Recommendations for usage of य and व

Closely associated with the usage of ऐ and औ cf. Section 3.10, one of the major source of errors is where to use the vowel letter and where to use the glides य and व .

Clear guidelines are laid down. Thus given किए and किये, the former i.e. the vowel letter is to be preferred over the latter. The same applies to ई and यी, the former being the preferred choice. This applies as the guidelines explicitly state to cases where:

जहाँ श्रुतिमूलक य, व का प्रयोग विकल्प से होता है, वहाँ इनका प्रयोग न किया जाए

However where

य श्रुतिमूलक व्याकरणिक परिवर्तन न होकर शब्द का ही मूल तत्व हो, वहाँ वैकल्पिक श्रुतिमूलक स्वरात्मक परिवर्तन करने की आवश्यकता नहीं है।

e.g. स्थायी, दायित्व etc.

जहाँ श्रुतिमूलक य, व का प्रयोग विकल्प से होता है वहाँ इनका प्रयोग न किया जाए, अर्थात्, किए : किये, नई : नयी, हुआ : हुवा आदि में से पहले (स्वरात्मक) रूपों का प्रयोग किया जाए। यह नियम क्रिया, विशेषण, अव्यय आदि सभी रूपों और स्थितियों में लागू माना जाए। जैसे :- दिखाए गए, राम के लिए, पुस्तक लिए हुए, नई दिल्ली आदि।

जहाँ 'य' श्रुतिमूलक व्याकरणिक परिवर्तन न होकर शब्द का ही मूल तत्व हो वहाँ वैकल्पिक श्रुतिमूलक स्वरात्मक परिवर्तन करने की आवश्यकता नहीं है। जैसे :- स्थायी, अव्ययीभाव, दायित्व आदि (अर्थात् यहाँ स्थाई, अव्यईभाव, दाइत्व नहीं लिखा जाएगा)।

3.2 LOAN-WORDS AND THEIR CORRECT REPRESENTATION

A large number of words have entered into the Hindi language and rules for their representation are important. The Central Hindi Directorate has studied this issue at length and provided rules and recommendations. These can be classified as under:

a. Urdu Loan words

Because of the close interaction between the two languages, these occupy the largest set of recommendations. The recommendations are as under:

Recommendations for Urdu words assimilated into Hindi:

Recommendation 1:

उर्दू से आए अरबी-फ़ारसी मूलक वे शब्द जो हिंदी के अंग बन चुके हैं और जिनकी विदेशी ध्वनियों का हिंदी ध्वनियों में रूपांतर हो चुका है, हिंदी रूप में ही स्वीकार किए जा सकते हैं। जैसे- कलम, किला, दाग आदि (क़लम, क़िला, दाग् नहीं)। पर जहाँ उनका शुद्ध विदेशी रूप में प्रयोग अभीष्ट हो अथवा उच्चारणगत भेद बताना आवश्यक हो, वहाँ उनके हिंदी में प्रचलित रूपों में यथास्थान नुक्ते लगाए जाएँ।

जैसे :- खाना : खाना, राज : राज, फन : हाइफ़न आदि।

Recommendation 2: The following words shall not be written as conjoined:

उर्दू मूल के निम्नलिखित शब्दों में संयुक्ताक्षर का प्रयोग न हो -

नुकसान, इनसान, इनकार, परवाह, बरबाद, मजदूर, सरकार, फरमान, तनखाह, करबट,
कुदरत, शिरकत, मरकज़, अकबर, मुजरिम, परवाना, मुगदर, मसखरा, इलजाम, फिलहाल,
फुरसत, मनज़र, (दृश्य), इनकलाब, आसमान, तकलीफ, मरहम, दरकार, शबनम,
इजलास, इकबाल, इनतहा, तनहा, तनहाई, असल, असली, असलियत, गलत, गलती,
नकल, नकली, नकलची, वरना, शरबत, गरदन, दफ्तर, सरकस, बरतन, बिलकुल, अकसर,
रफ़तार, गरम, गरमी, शरम, शरमाना, सरदी आदि। (इनसान, इनकार, तनखाह, मनज़र,
इनकलाब, इनतहा, तनहा, तनहाई आदि को किसी भी स्थिति में इंसान, इंकार, तंखाह,
मंज़र, इंकलाब, इंतहा, तंहा, तंहाई के रूप में नहीं लिखा जाना चाहिए)।

Recommendation 3: However in the case of the following words, these will be obligatorily conjoined:

फुर्ती, फुर्तीला, शिकस्त, दुरुस्त, बुजुर्ग, गिरफ्तार, रुख्सत, ज़िंदा, कर्ज़, खर्च, ज़िंदगी, इंतकाम, इंतज़ार, इल्तजा, अशफौ, इम्तहान, तिलिस्म, पुछता, दरखत, तरार, तश्तरी, लश्कर, जिन्स, मुंशी, नकशा, कब्ज़, लफ़्ज़, शुक्र, शुक्रिया अखियार, इत्तेफाक, इश्तहार, बख्शीश, अब्ल, शब्ल, सख्त, सख्ती, गश्त, किस्त, किस्म, मुश्क, इश्क, जश्न, कश्मीर, पश्मीना, जुल्म, जुर्म, जिस्म, रस्म, रस्मी, हश्र, फिक्र, मुल्क, इल्म, फर्ज, फर्जी, ईर्द-गिर्द, गोश्त, रास्ता, नाश्ता, शख्स, दर्ज, दर्जी, सुख्ख, सुख्खी, गुर्दा, बर्फ, जल्द, जल्दी, चश्मा, भिश्ती, सर्द, उम्र, कुश्ती, चुस्त, चुस्की, फर्क, शर्म, शर्मीली, बर्फी आदि।

Some words admit both forms as shown below. However in this case, the first form shall be treated as standard and the second form be not used

हिंदी में कुछ प्रचलित शब्द ऐसे हैं जिनकी वर्तनी के दो-दो रूप बराबर चल रहे हैं। विद्वत्समाज में दोनों रूपों की एक-सी मान्यता है। कुछ उदाहरण हैं :- गरदन/गर्दन, गरमी/गर्मी, बरफ़/बर्फ़, बिलकुल/बिल्कुल, सरदी/सर्दी, कुरसी/कुर्सी, भरती/भर्ती, फुरसत/फुर्सत, बरदाश्त/बर्दाश्त, बरतन/बर्तन, दुबारा/दोबारा आदि। इन वैकल्पिक रूपों में से पहले वाले रूप को प्राथमिकता दी जाए।

Recommendation 4: The following standard shall be adopted in the case of Hindi as well as Urdu words:

हिंदी तथा उर्दू के निम्नलिखित शब्द मानक होंगे -

दुहरा, दुहराना, दुगुना, दूना, दुपहर, दुमज़िला, दुबारा, दुधारी, दुकान, दुशाला, कुहनी, कुहरा, कुहराम, मुहब्बत, कुहासा, सुहाग, सुहागन, मोहताज, शोहरत, बोहनी, मोहसिन, पहनावा, पहचान, दहलीज, मेहमान, एहसान, एहसास, एहतियात, महँगा, लहँगा, तहज़ीब, गहरा, मुहल्ला, धुआँ, कुआँ, कुँवर, कुँआरा, गेहुँआ आदि।

Recommendation 5: Urdu words may be adopted in Hindi but under the following proviso:

हिंदी में उदू के सभी शब्द आवश्यकतानुसार अपनाए जा सकते हैं। ये हिंदी के लिए विदेशी शब्द नहीं माने जाएँगे। हाँ, आवश्यकतानुसार उनका कोष्ठकों में अर्थ देना उपयुक्त है। जैसे :- दौरान (अवधि में), की निस्बत (के लिए) आदि।

मुख्यालफ़त (विरोध), ज़ाहिर (स्पष्ट), हमला (आक्रमण), कामयाबी (सफलता), मुहैया कराना (उपलब्ध कराना), पेश करना (प्रस्तुत करना), मक्सद (उद्देश्य), हद (सीमा), जिहादी (धर्म युद्ध करने वाला), शुरुआत (आरंभ), कोताही (लापरवाही), गैरतलब (ध्यान देने योग्य), हैरत (आश्चर्य), उस्ताद (अध्यापक), कानूनन (विधि के अनुसार), अदालत (न्यायालय), अगरचे (यद्यपि) आदि।

दरअसल = असल में ('दरअसल में' नहीं)

दौरान = अवधि में या दौर में ('दौरान में' नहीं)

क्रौम = बहुवचन क्रौमें ('अक्वाम' नहीं)

काग़ज़ = बहुवचन काग़ज़ों (में) या काग़ज़ात में (काग़ज़ातों में नहीं)

किताब = बहुवचन किताबें (किताबात नहीं)

किस्म = किस्में (अक्साम नहीं)

b. Loans from Indian languages

These cover languages as varied as Marathi, Bangla, Punjabi, Odia, Santali, Gujarati and Dravidian languages. Some examples of such loans are provided below:

मराठी शब्द - बाजू, लागू, चालू आदि।

बांगला शब्द - रसगुल्ला, चमचम, उपन्यास, गल्प, आपत्ति आदि।

पंजाबी शब्द - सिक्ख, खालसा, भांगड़ा आदि।

ओड़िआ शब्द - अटका आदि।

द्रविड़ परिवार की भाषाओं के शब्द- डोसा, इटली, मीन, पिल्ला, सांभर आदि।

संताली शब्द - कोड़ी, गण्डा, गाही आदि।

गुजराती शब्द - हડ़ताल, गरबा आदि।

In the case of Sanskrit, the Central Hindi Directorate prescribes the following:

संस्कृत के तत्सम शब्द ही मानक होंगे लेकिन परबत (पर्वत), दरपन (दर्पण) जैसे शब्दों को कोशकार 'उर्दू शैली', 'बोली' आदि नामांकन से कोशों में प्रविष्ट करेंगे।

- c. **Loans from Foreign Languages:** These cover the following languages: English , Arabic, Persian, Turkish, Portuguese, French, Spanish, German, Japanese, Russian, Italian. Guidelines as to their usage are provided below:

Rule 1: Those words assimilated in Hindi and shown in Hindi dictionaries will be treated as such:

अंग्रेज़ी - बटन, फ़ीस, पिन, पेट्रोल, पुलिस, पैसिल, बूट, निब, कैमरा, फ़ोटो, रेडियो, सिनेमा, साइकिल, कैलेंडर, फ्रेम, पैन्ट, जींस, लेंस, नाइट्रोजन, डॉक्टर, कॉलेज, ऑफिस, कॉल, कॉन्फ्रेंस, टॉफ़ी, टॉकीज़, पॉलिश, फॉर्म, फुटबॉल, मनीऑर्डर, हॉल, ऑक्सीजन आदि।

अरबी - (प्रायः फारसी के माध्यम से) कब्र, ख़राब, काग़ज़, कानून, अल्लाह आदि।

फ़ारसी- कमर, कम, ख़ाक, गुम, वापस, खुदा आदि।

तुर्की - चाकू, तोप, लाश, उर्दू (छावनी), जैसे :- उर्दू बाज़ार = छावनी, मोरी (घोड़ा), उदाहरण- मोरी गेट, तमगा, चकमक आदि।

पुर्तगाली- (अधिकतर गुजराती, मराठी के माध्यम से) अलमारी, कमीज़, कमरा, मेज़, इस्पात, तौलिया, नीलाम, कमरा, पादरी, काजू आदि।

फ्रांसीसी- कारतूस, अंग्रेज़ (<आंग्लिस), ऐडवोकेट, मेयर, लैंप, वारंट, सूप, कूपन आदि।

स्पेनी- सिगार, सिगरेट, कार्क, पितन आदि।

जर्मन- ट्रेन, सेमिनार आदि।

जापानी- रिक्षा, जूड़ो आदि।

रूसी- सोवियत, स्पुतनिक, जार आदि।

इतालियन- लॉटरी, कारटून, रॉकेट, गजट, सॉनेट, मलेरिया, स्टूडियो, पिआनो, वायलिन आदि।

Rule 2:-In the case of other words major guidelines are given below:

- The generic rule is as under. All foreign words will be indicated by placing them in [‘ ’]

अन्य सारे विदेशी शब्दों को हिंदी में प्रयुक्त करना हो तो उन्हें शब्द चिह्न (सिंगल इन्वर्टेड कॉमा) [‘ ’] में लिखना होगा। जैसे :- मैं जानता हूँ कि यह एक ‘प्रॉब्लम’ है।

- In the case of English words, only the long vowels ई ऊ shall be used, as illustrated below:

अंग्रेजी शब्दों का लिप्तंतरण करते समय शब्दांत में दीर्घ 'ई' तथा 'ऊ' का ही प्रयोग होगा। जैसे :-

dye	डाइ	waterloo	वाटरलू
pie	पाई	shoe	शू
city	सिटी	pharmacy	फार्मसी
daddy	डैडी	company	कंपनी
luxury	लग्जरी	blue	ब्लू ... आदि।
key	की		

- Insofar as words of Latin and Slavic origins are concerned the /t/ will be mapped to /त/ as in the examples given below.

लातीनी और स्लाव परिवार की भाषाओं (फ्रांसीसी, स्पेनी, रूसी) तथा जापानी आदि 'त' वर्गीय तथा कुछ अन्य भाषाओं के नामों के उच्चारण को सुरक्षित रखा जा सके तो अच्छा होगा। जैसे :-

फ्रांसीसी	मितराँ (Mitterand)	रूसी	तोल्स्तोय/तोल्स्तोई (Tolstoy)
दूप्ले (Duplex)		स्पेनी	दान किहोते (डॉनकिवग्ज़ोट)
		जापानी	तोक्यो, क्योतो

- A further guideline specifies that in the case of proper names, these be transcribed as such and not translated as shown below. This rule has repercussions for CLDR where very often Place Names, Names of Languages i.e. Named Entities are translated and not transliterated as prescribed by the Central Hindi Directorate.

अन्य	पोल्स्की ('पोलिश' नहीं)
	सोल ('सियोल' नहीं)
	म्याँमा ('म्याँमार', 'मियाँमार' नहीं) ... आदि।

4. PUNCTUATIONS¹⁸

4.1. This section provides recommendations for correct use of punctuation markers. Punctuation in Indian languages is a topic that has been considerably ignored and word-processing softwares do not apply punctuation rules. The Central Hindi Directorate lays down a clear style guide for use of Punctuations and developers would do well to follow the style guide.

All the punctuation markers used in Hindi are studied and clear rules and norms are prescribed for their use. Chapter 4 of the document treats of punctuation at length.

This section reproduces the list of Punctuation markers along with their Hindi equivalents and their Unicode values already provided in 6.3.1.9. of the DEVANAGARI SCRIPT BEHAVIOUR FOR HINDI:

Devanagari Name	Name of the marker	Marker Shape
"स्टॉप" ¹⁹	U+002E FULL STOP	.
प्रश्न चिह्न	U+003F QUESTION MARK	?
अल्पविराम	U+002C COMMA	,
विस्मयसूचक चिह्न	U+0021 EXCLAMATION MARK	!
ऊर्ध्व अल्पविराम	U+0027 APOSTROPHE	,
अर्धविराम/ सेमीकोलन	U+003B SEMICOLON	;
उपविराम	U+003A COLON	:
योजक चिह्न	U+002D HYPHEN-MINUS	-
निर्देशक चिह्न	U+002D HYPHEN-MINUS U+002D HYPHEN-MINUS Dash	--
लोप चिह्न	FULL STOP (U+002E) FULL STOP (U+002E) FULL STOP (U+002E) Elimination Sign	...
	U+002F SLASH	/
उद्धरण चिह्न	U+0022 QUOTATION MARK	" "
शब्द चिह्न	U+0027 APOSTROPHE SINGLE QUOTATION MARK	„
	U+0058 LATIN CAPITAL LETTER X used as a cross-out character.	X

¹⁸ Section 4.0 of the document. The correct Unicode values of punctuations are provided

¹⁹ <http://hi.wikipedia.org/s/ys0>

	U+002D HYPHEN-MINUS U+002D HYPHEN-MINUS U+0022 QUOTATION MARK U+002D HYPHEN-MINUS U+002D HYPHEN-MINUS as above	--"--
कोष्ठक	U+0028 LEFT PARENTHESIS U+0029 RIGHT PARENTHESIS	()
कोष्ठक	U+005B LEFT SQUARE BRACKET U+005D RIGHT SQUARE BRACKET	[]
कोष्ठक	U+007B LEFT CURLY BRACKET U+007D RIGHT CURLY BRACKET	{ }
संक्षेपसूचक चिह्न	U+0970 DEVANAGARI ABBREVIATION SIGN	◦
पूरण विराम	U+0964 DEVANAGARI DANDA Devanagari Purna Viram	
दीर्घ विराम	U+0965 DEVANAGARI DOUBLE DANDA Devanagari Deergha Viram	
हंसपद त्रुटिबोधक	/ Sign of left/Add word	↖

4.1 Full Stop: (।) or पूर्ण विराम²⁰

4.1.1 वाक्य के अंत होने की सूचना। जैसे- श्याम पढ़ता है।

टिप्पणी : प्रश्नवाचक तथा विस्मयसूचक वाक्यों में पूर्णता को प्रकट करने वाले चिह्न क्रमशः :- ? तथा ! हैं।

4.1.2 अप्रत्यक्ष प्रश्न के अंत में भी पूर्ण विराम लगता है। जैसे :-

आपने बताया नहीं कि आप कहाँ जा रहे हैं।

आप कहाँ जा रहे हैं, यह आपने बताया नहीं।

कृपया सूचित कीजिए कि आप यहाँ कब आ रहे हैं।

4.1.2.1 अगर मूल उपवाक्य ही प्रश्न हो तो अंत में प्रश्न चिह्न आएगा। जैसे :-

क्या आप बता सकते हैं कि वे कहाँ जा रहे हैं?

4.1.3 शब्द, पद तथा उपशीषकों के अंत में पूर्ण विराम न दिया जाए। जैसे :-

प्राणी तीन तरह के हैं : (क) थलचर (ख) जलचर (ग) नभचर

²⁰ The term as used by the Central Hindi Directorate refers to the purna viram and not the English full stop.

4.2 Interrogation Marker: (?)

प्रश्न पूछने की स्थिति में पूर्ण विराम के स्थान पर प्रश्नवाचक चिह्न लगाया जाता है। इसके प्रयोग की स्थितियाँ निम्नलिखित हैं-

- 4.2.1 प्रश्नवाचक वाक्यों के अंत में। जैसे :-
तुम्हारा नाम क्या है? वह कहाँ रहता है?
- 4.2.2 यदि एक वाक्य में कई प्रश्नवाचक उपवाक्य हों तो पूरे वाक्य के समाप्त होने पर, सबसे अंत में। जैसे :-
मैं क्या करता हूँ, मैं कहाँ-कहाँ जाता हूँ, कहाँ क्या खाता हूँ, यह सब जानने के आप इच्छुक क्यों हैं?
- 4.2.3 एक साथ जब कई प्रश्न हों तो प्रश्न चिह्न पृथक्-पृथक् भी लगाया जाता है।
जैसे :- कहाँ? किसे? किस समय? किसने साक्षर किया?
- 4.2.4 प्रश्नयुक्त मुहावरे के अंत में। जैसे :-
जंगल में मोर नाचा किसने देखा ?

4.3 Exclamation (!):

The rules for the use of exclamation markers are laid down below.

4.3.1 हर्ष, विषाद, धृणा, आश्चर्य, विस्मय आदि प्रकट करने वाले शब्दों, उपवाक्यों अथवा वाक्य के अंत में -

कितना सुहावना है यह स्थान! हाय! बेचारा मारा गया!
इतनी लंबी दीवार! ईश्वर न करे!
चिरंजीवी हो!

4.3.2 तीव्र मनोविकार व्यक्त करने के लिए -

4.3.2.1 गधा कहीं का! निपट गँवार!

4.3.2.2 निरंतर वृद्धि के लिए दो चिह्न, फिर तीन चिह्न -
शोक! शोक!! महाशोक!!!

4.3.3 प्रश्नवाचक वाक्य के अंत में भी तीव्र मनोवेग प्रदर्शित करने के लिए -
बोलते क्यों नहीं, क्या गूँगे हो!

4.3.4 संबोधन के लिए -

मोहन! बोलो, क्या कह रहे थे?
बोलो मोहन! क्या कह रहे थे?
बेटे! तुम ठीक कह रहे थे।

मित्रो! आज सब कुछ त्याग देने का समय आ गया है।
ओ राम! देखो, इधर देखो।

4.3.5 स्वीकृति/अस्वीकृति प्रकट करने के लिए -

अच्छा! पर लंबा चक्कर मत लगाना।
यह हरगिज़ नहीं हो सकता!

4.3.6 अविश्वसनीयता प्रकट करने के लिए -

क्या कहा! वह पकड़ा गया?

4.4 Colon: (:)

Colon should not be confused with Visarga. As a punctuation marker it has a wide variety of uses, all of which are laid down below.

4.4.1 सामान्यतः इसका प्रयोग आगे आने वाली सूची आदि के पूर्व किया जाता है।
जैसे - महत्वाकांक्षी के तीन शात्रु हैं : आलस्य, हीन-भावना एवं पराश्रय।

4.4.2 संख्याओं में अनुपात प्रदर्शित करने के लिए -

1 : 2 5 : 6 7 : 8 9 : 10

4.4.3 नाटक/एकांकी में उद्धरण चिह्नों के प्रयोग के बिना मात्र कोलन का प्रयोग किया जाता है -

सोना : राम की कसम देखिए!

4.4 स्थान, समय की सूचना के लिए -

स्थान : समिति कक्ष

समय : अपराह्न चार बजे

8 : 32 : 10 = आठ बजकर बत्तीस मिनट और दस सेकंड

प्रातः 7 : 30 = प्रातः सात बजकर तीस मिनट

4.5 Semi-colon: (;

The semi-colon should be used in the following cases in preference to the comma.

अर्धविराम (;) उपविराम तथा अल्पविराम के मध्य की स्थिति है। अंग्रेज़ी में इसको सेमीकोलन की संज्ञा दी गई है जिसकी आकृति कोलन के बिंदु तथा कॉमा के योग से बनाई गई है। इसके प्रयोग के कुछ नियम निम्न प्रकार हैं -

- 4.5.1 छोटे-छोटे दो से अधिक वाक्यों की शृंखला में, जो एक ही वाक्य पर आश्रित हों -
विद्या विनय से आती है; विनय से पात्रता; पात्रता से धन; धन से धर्म और धर्म से सुख की प्राप्ति होती है।
- 4.5.2 समुच्चयबोधकों से बने उपवाक्यों को पृथक् करने के लिए इस चिह्न का प्रयोग किया जाता है -
आपने उसकी निंदा की; अतएव वह आपका दुश्मन ही बनेगा।
स्वाभिमान मनुष्य को सहारा देता है; फिर भी कितने हैं जो इसको अपनाते हैं।
- 4.5.3 समानाधिकरण उपवाक्यों के बीच में -
भारत में राजनीतिक स्वतंत्रता का शंख महात्मा गांधी ने फूँका; अहिंसा और असहयोग के अस्त्र उन्होंने दिए।
सूर्य पृथ्वी से बहुत दूर है; वह पृथ्वी से काफी बड़ा है।
- 4.5.4 अंकों के विवरण देने में -
भारत - 210; पाकिस्तान - 188; जापान - 205

4.6 Comma (,)

- 4.6.1 वाक्य, वाक्यांश में तीन या उससे अधिक शब्दों को अलग करने के लिए जैसे :-
दिल्ली, मुंबई, कोलकाता और चेन्नै भारत के बड़े नगर हैं।
वह तो अपनी भूमि, संपत्ति, प्रतिष्ठा और मान-मर्यादा सभी कुछ खो बैठा।
- 4.6.2 एक ही प्रकार के कई पदबंधों, उपवाक्यों का जब एक वाक्य में प्रयोग हो तो उनको अलग करने के लिए जैसे :-
केरल की मोहक झीलें, सुंदर बाग और प्राकृतिक सौंदर्य सबको मोह लेते हैं।
वह प्रतिदिन आता है, काम करता है और लौट जाता है।
- टिप्पणी : अंतिम शब्द, पदबंध, उपवाक्य से पहले अल्पविराम का प्रयोग न कर 'और' का प्रयोग किया जाता है।
- 4.6.3 वाक्य के मध्य में किसी क्षिप्त वाक्यांश/उपवाक्य को अलग दिखाने के लिए -
गणित का पाठ्यक्रम बदल जाने से, मैं समझता हूँ, इस वर्ष हायर सैकेंडरी का परीक्षाफल प्रभावित होगा।
- 4.6.4 सकारात्मक और नकारात्मक शब्दों में 'हाँ' या 'नहीं' के बाद -
हाँ, मैं तो जरूर आऊँगा।
नहीं, मैं नहीं आऊँगा।
- 4.6.5 बस, सचमुच, अच्छा, वास्तव में, खैर आदि से प्रारंभ होने वाले वाक्यों में इनके बाद -
बस, देख लिया तुम्हें।
सचमुच, तुम बड़े अच्छे हो।
[तुलना करें - तुम सचमुच बड़े अच्छे हो।]
अच्छा, तो लीजिए और चलते बनिए।
वास्तव में, उसकी कार्य-पद्धति प्रशंसनीय है।
खैर, जो हुआ सो अच्छा!
- 4.6.6 उद्धरण से पूर्व -
हरिमोहन ने कहा, "मैं इस बार चुनाव में खड़ा हो रहा हूँ।"

- 4.6.7 संबोधन के बाद -
बेटा, सवेरे जल्दी उठकर पढ़ने की आदत डालो।
- 4.6.8 जो उपवाक्य किंतु, लेकिन, क्योंकि, पर, परंतु, अतः आदि समुच्चयबोधक से प्रारंभ होते हैं, उनमें समुच्चयबोधक अव्यय से पहले अल्पविराम (,) लगाया जाता है। जैसे :-
मैं तुम्हारे साथ नहीं खेलूँगा, किंतु तुम्हारा खेल ज़रूर देखूँगा।
वह मेरे ही होस्टल में रहता है, लेकिन एक बार भी मुझसे नहीं मिला।
मैं आज कार्यालय नहीं आऊँगा, क्योंकि मेरे घर मेहमान आए हुए हैं।
उन्होंने पहले तो आने का वादा किया, पर बाद में मुकर गए।
माँ घर में रहती हैं, अतः वह बच्चों की देखभाल कर लेती हैं।
- 4.6.9 भावातिरेक में किसी शब्द/शब्दसमूह पर बल देने के लिए, जब पुनरावृत्ति हो -
दौड़ो, दौड़ो, बम फट गया।
- 4.6.10 शोक की अभिव्यक्ति, विस्मयादिबोधक शब्दों के बाद -
हाय, मैं तो लुट गई।
धिक्कार है, तुमने ऐसा ओछा काम किया।
- 4.6.11 कभी-कभी तब, वह, तो, कि आदि संयोजक शब्दों के स्थान पर अल्पविराम का प्रयोग किया जाता है। जैसे :-
जब स्टेशन पहुँचे, पानी तेज़ हो गया। (तब)
जो टिप्पणी कल मंत्रीजी ने भेजी है, कहाँ है? (वह)
जब हमें देर हो गई, हम उसके घर ही रुक गए। (तो)
कहते हैं, आज से छह हज़ार साल पहले महाभारत का युद्ध हुआ था। (कि)

- 4.6.12 समानाधिकरण शब्द/पदबंध के मध्य में इसका प्रयोग अवश्य किया जाना चाहिए। यह प्रवृत्ति अंग्रेज़ी के प्रभाव से बढ़ी है, पर कुछ स्थितियों में नितांत आवश्यक हो गई है। जैसे :-
मैं, अशोक शर्मा, यह घोषणा करता हूँ कि
- 4.6.13 तारीख के साथ माह का नाम लिखने के बाद -
15 अगस्त, 1947 को भारत स्वतंत्र हुआ।
- 4.6.14 पत्र में संबोधन के बाद -
महोदय,
प्रिय महोदय,
- 4.6.15 जब विशेषण उपवाक्य मध्य में ढाल दिया जाए -
वह रोगी, जिसे कल अस्पताल में दाखिल किया था, आज ठीक हो गया।
मेरा भाई, जो डॉक्टर है, कल अमेरिका गया।
टिप्पणी : यह अंग्रेज़ी की प्रकृति है। इसके स्थान पर हिंदी की प्रकृति के अनुकूल लिखना उचित है।
- 4.6.16 संख्याओं को पृथक् करने के लिए -

1, 25, 525

12, 20, 424

4.7 Apostrophe (‘)

Apostrophe is mainly used where the English “apostrophe” is called for, as in dates

- 4.7.0 ऊर्ध्व अल्पविराम (') का प्रयोग अंग्रेजी के अपॉस्ट्रॉफी के लिए किया जाता है। अंग्रेजी में इस चिह्न का प्रयोग संबंधकारक का भी सूचक है, पर हिंदी में तो यह चिह्न मात्र ‘अंक लोप’ का ही सूचक है। जैसे :-
सन् '47 से सन् '50 तक = सन् 1947 से सन् 1950 तक

4.8 Dash (-)

The dash sign needs to be clearly distinguished from the hyphen sign (cf. 4.17 below). This punctuation marker is used in the following cases:

- 4.8.0 निर्देशक चिह्न (-) का आकार योजक चिह्न (-) से अधिक लंबा होता है। इसको रेखिका भी कहते हैं। निम्नलिखित मुख्य-मुख्य स्थितियों में इसका प्रयोग किया जाता है -
- 4.8.1 किसी व्यक्ति के वाक्य को उद्धृत करने से पूर्व -
अध्यापक - भारत के प्रथम राष्ट्रपति कौन थे?
- 4.8.2 कहना, लिखना, बोलना, बताना आदि क्रियाओं के बाद -
कमला ने कहा - मैं कल चली जाऊँगी।
- 4.8.3 'निम्नलिखित' आदि शब्दों के बाद -
नाम निम्नलिखित हैं - सीता, भारती, विमला।
परिश्रम से सब-कुछ प्राप्त हो सकता है - सुख, संपत्ति, यश, प्रतिष्ठा।
- 4.8.4 किसी अवतरण के बाद और उसके लेखक/वक्ता से पहले -
स्वराज्य हमारा जन्मसिद्ध अधिकार है - तिलक।
- 4.8.5 क्षिप्त वाक्यांशों के आगे और पीछे -
लिखना - मौलिक रचना करते रहना - मेरा काम है। (यहाँ दोनों स्थानों पर अल्प विराम का विकल्प भी है) (पूर्व संदर्भ 4.6.3)।
- 4.8.6 किसी वाक्य के अंतिम खंड पर ज़ोर देने के लिए -
हाँ, हाँ, आप आइए - बड़े शौक से।
- 4.8.7 समानाधिकरण शब्दों को पृथक् करने के लिए -
सभी लोगों ने - श्रमिकों, बुद्धिजीवियों ने इस परिवर्तन का स्वागत किया।
(इसके स्थान पर कोष्ठक () का प्रयोग भी होता है।)
- 4.8.8 संख्याओं, सन् या दो नामों के बीच इसका प्रयोग किया जाता है, तब उसका अर्थ होता है - 'अमुक' से 'अमुक' तक।
पृष्ठ 40 - 55 यानी पृष्ठ 40 से 55 तक।
सन् 1955 - 60 (सन् 55 से सन् 60 तक)
चंडीगढ़ - दिल्ली (चंडीगढ़ से दिल्ली तक)

- 4.8.9 व्याख्या करने के लिए। जैसे -
जिस तरह हो सके - साम, दाम, दंड, भेद - तुम्हें यह काम करना ही होगा।
- 4.8.10 'कि' के स्थान पर। जैसे -
होरी ने कहा - तो तू क्या समझती है, मैं बूढ़ा हो गया?
- 4.8.11 कोश में शब्द के विवरण के अंतर्गत अन्य शब्दों के निर्माण के लिए मुख्य शब्द की सूचना देने के लिए [-] का प्रयोग किया जाता है।
'प्रयोग' के अंतर्गत,
- कर्ता
- धर्मी
- शाला
(प्रयोगकर्ता, प्रयोगधर्मी, प्रयोगशाला)

4.9 Colon Dash (:-)

यह चिह्न कोलन/उपविराम और निर्देशक चिह्न/रेखिका का मिश्रित रूप है। अतः इन दोनों ही अलग-अलग चिह्नों के स्थान पर वैकल्पिक रूप में इस चिह्न का प्रयोग भी प्रायः किया जाता है। जैसे :-

किसी एक कवि का जीवन परिचय लिखिए :- 1. तुलसीदास 2. कबीरदास
3. सेनापति

4.10 Inverted comma (“ ”)

- 4.10.0 किसी का कहा वाक्य ज्यों-का-ज्यों उद्धृत करते समय, नाटकों में संवाद के साथ तथा किसी सिद्धांत वाक्य अथवा लोकोक्ति के साथ इस चिह्न का प्रयोग करने की प्रथा है।
- 4.10.1 कथन का उद्धरण -
सुभाषचंद्र बोस ने कहा था, “तुम मुझे खून दो, मैं तुम्हें आज़ादी दूँगा।” (वक्ता का उल्लेख पहले, कथन बाद में)
“क्यों” उसने कुछ विस्मय से पूछा। (कथन पहले, वक्ता का उल्लेख बाद में)
- 4.10.2 नाटकों में संवाद के लिए -
यवन - “क्यों, गांधार नरेश ने तुम्हें क्या आज्ञा दी है?”
सैनिक - “यही कि आप जिसे कहें उसे हम लोग बंदी बनाकर महाराज के पास ले चलें।”
- 4.10.3 सिद्धांत वाक्य या लोकोक्ति -
“मा फलेषु कदाचन”
“थोथा चना बाजे घना”
टिप्पणी : नाटक के संवादों में प्रायः उद्धरण चिह्नों को छोड़ दिया जाता है। बिना उद्धरण चिह्नों के भी संवाद के कथन चिह्न युक्त ही समझे जाएँगे।

4.11 Single Inverted Comma: (‘ ’)

- 4.11.1 किसी पुस्तक, व्यक्ति आदि के नाम को एक कॉमा/अल्पविराम जैसे उद्धरण चिह्न में रखा जाता है। जैसे –
‘बालभारती’ बच्चों की पत्रिका है।
दूसरा अध्याय पढ़ो, जिसका शीर्षक है – ‘बात’।
- 4.11.2 किसी शब्द/पद को पृथक् रूप से उभारने के लिए –
“वह पुस्तक लिखता है।” इस वाक्य में ‘लिखता है’ क्रिया का कर्म ‘पुस्तक’ है।
“वह अच्छा लड़का है।” इस वाक्य में ‘अच्छा’ शब्द ‘लड़का’ की विशेषता बता रहा है।
- 4.11.3 व्याख्या के लिए –
इस भाषा का नाम ‘संस्कृत’ पड़ा, जिसका शाब्दिक अर्थ है ‘सुधरा/सुधारा हुआ’।

4.12 Brackets: [{()}]

- 4.12.1 छोटे कोष्ठक () के भीतर मुख्यतः उस सामग्री को रखते हैं, जो मुख्य वाक्य का अंग होते हुए भी पृथक् की जा सकती है। जैसे :-
संज्ञा के तीन मुख्य भेदों (व्यक्तिवाचक, जातिवाचक, भाववाचक) का विवेचन किया जा रहा है।
- 4.12.2 क्रमसूचक अंकों या अक्षरों के साथ -
(1) (2) (3) (क) (ख) (ग)
- 4.12.3 शब्दों या अंकों में लिखी संख्या को लिखने के लिए भी कोष्ठक का प्रयोग किया जाता है। जैसे :-
(क) पाँच सौ रुपए (₹ *500/-)
(ख) ₹ 240/- (₹ दो सौ चालीस मात्र)
- 4.12.4 लेखक या कृति का नाम सूचित करने के लिए : 'कामायनी' (जयशंकर 'प्रसाद'); जयशंकर 'प्रसाद' (कामायनी); "मा फलेषु कदाचन" (गीता)
- 4.12.5 संदर्भ या निर्देश के रूप में प्रस्तुत सामग्री -
प्राचीन भारत में राजा निरंकुश नहीं होता था। (देखिए : अध्याय पाँच)
- 4.12.6 नाटक में रंगमंच संबंधी निर्देशों के लिए -
महेंद्र (शून्य में देखते हुए) - आश्चर्य!
आरोपी (शार्तिपूर्वक) - "नहीं, नहीं!"
- 4.12.7 भाषण, वाद-विवाद, पठन आदि में यदि कोई टिप्पणी देनी हो -
नेहरू - "आराम हराम है " (तालियाँ)
- 4.12.8 समाचार पत्रों में प्रकाशित समाचारों के प्रारंभ में ही यह सूचना कोष्ठक में दी जाती है कि अमुक समाचार किस समाचार एजेंसी के माध्यम से प्राप्त हुआ है।
जैसे :-
नई दिल्ली, 18 मार्च (भाषा) - पी.टी.आई. की हिंदी सेवा 'भाषा' के नाम से व्यक्त की जाती है।
बहरीन, 18 मार्च (प्रेट्र) - प्रेट्र = प्रेस ट्रस्ट
वाराणसी, 18 मार्च (वार्ता) - यू.एन.आई. की हिंदी सेवा 'वार्ता' के नाम से जानी की जाती है।

4.12.9 किसी भी सामग्री के अंत में नोट/टिप्पणी कोष्ठक में दी जाती है। जैसे -
विज्ञापन
आवश्यकता है
सूचना

(नोट : केवल दिल्ली निवासी ही आवेदन करें। सोमवार को कार्यालय बंद रहता है।)

टिप्पणी : छोटे कोष्ठक के अलावा मँझला कोष्ठक/सर्पाकार कोष्ठक { } तथा
बड़ा कोष्ठक [] भी होता है जब किसी कारणवश दो या तीन प्रकार

की टिप्पणियाँ/सूचनाएँ देनी हों तो छोटा कोष्ठक मँझले कोष्ठक में समा
जाता है, और ये दोनों कोष्ठक मिलकर फिर बड़े कोष्ठक में समाहित
कर दिए जाते हैं।

4.13 Elimination Sign: (...)

4.13.1 लोप चिह्न मुख्य रूप से तीन छोटे-छोटे बिंदुओं से प्रदर्शित किया जाता है। जब किसी शब्द का जान बूझकर प्रयोग न करना हो या तत्काल याद न आ रहा हो तो लोप चिह्न प्रयुक्त होता है और जब वात छूटी हुई हो तो तीन-तीन बिंदुओं को कई बार लगाया जाता है। जैसे :-

तेरी ... तैसी।

अरे हाँ ... अब याद आ गया।

यदि आप ... चाहें तो

4.13.2 किसी गद्यांश का उल्लेख आवश्यक न हो या कविता के मध्य का भाग छोड़ना हो तो लोप चिह्न का प्रयोग किया जाता है। जैसे :-

कवि कुछ ऐसी तान सुनाओ।

...

...

विश्वभर की पोषक वीणा

के सब तार मूक हो जाएँ

4.14 Abbreviation Sign (.) (◦)

संक्षिप्त बिंदु (◦) का प्रयोग शून्य (.) की भाँति भी होता है। जैसे :- एम.ए, ~ एम.ए. (मास्टर ऑफ आर्ट्स), रा.च. प्रसाद ~ रा.च. प्रसाद (राम चंद्र प्रसाद), प. ~ प. (पर्डित), से.मी. ~ से.मी. (सेंटीमीटर), दे. ~ दे. (देखिए) आदि।

4.15 Sign of left/Add word (↖)

लिखते समय जब कोई पद रह जाता है अथवा बाद में विचार आने पर जोड़ना होता है तो उसके लिए ठीक उस स्थान पर इस चिह्न का (↖)प्रयोग किया जाता है। वस्तुतः यह विराम चिह्न नहीं है। लेखन या प्रूफ रीडिंग में इसका प्रयोग किया जाता है। जैसे :-

हम समाजवादी के व्यवस्था हामी हैं।

वर्षा ऋतु में आकाश में बादल छा जाते हैं।

4.16 Hyphen: (-)

योजक चिह्न का प्रयोग इन स्थितियों में किया जाता है -

1. द्वंद्व समास के बीच में। जैसे :- माता-पिता, नर-नारी, भाई-बहिन राम-लक्ष्मण आदि।
2. दो समानार्थी शब्दों की पुनरुक्ति के बीच में। जैसे :- रोम-रोम, घर-घर, दूर-दूर आदि।
3. दो परस्पर विलोम शब्दों के बीच में। जैसे :- रात-दिन, ठंडा-गरम, उठना-बैठना आदि।
4. साम्य सूचक सा, से, सी से पूर्व में। जैसे :- छोटा-सा, मुझ-सा, हरी-सी आदि।
5. तत्पुरुष समास में समस्तपद से भ्रांति की संभावना होने पर। जैसे :- भू-तत्व।

5. CORRECT SPELLINGS OF NUMERALS:

The section on Numerals comprises the following parts:

a. Cardinals:

Since spellings of Cardinal Numbers is a major issue in all Indian languages, the following is provided as a norm to be followed.

एक	सोलह	इकतीस	छियालीस	इकसठ	छिहत्तर	इक्यानवे
दो	सत्रह	बत्तीस	सेंतालीस	बासठ	सतहत्तर	बानवे
तीन	अठारह	तैंतीस	अड़तालीस	तिरसठ	अठहत्तर	तिरानवे
चार	उन्नीस	चौंतीस	उनचास	चौंसठ	उनासी	चौरानवे
पाँच	बीस	पैंतीस	पचास	पैंसठ	अस्सी	पचानवे
छह	इक्कीस	छत्तीस	इक्यावन	छियासठ	इक्यासी	छियानवे
सात	बाईस	सैंतीस	बावन	सड़सठ	बयासी	सतानवे
आठ	तेईस	अड़तीस	तिरपन	अड़सठ	तिरासी	अठानवे
नौ	चौबीस	उनतालीस	चौवन	उनहत्तर	चौरासी	निन्यानवे
दस	पच्चीस	चालीस	पचपन	सत्तर	पचासी	सौ
ग्यारह	छब्बीस	इकतालीस	छप्पन	इकहत्तर	छियासी	
बारह	सत्ताईस	बयालीस	सत्तावन	बहत्तर	सतासी	
तेरह	अट्ठाईस	तैंतालीस	अट्ठावन	तिहत्तर	अठासी	
चौदह	उनतीस	चवालीस	उनसठ	चौहत्तर	नवासी	
पंद्रह	तीस	पैंतालीस	साठ	पचहत्तर	नब्बे	

हजार, लाख, करोड़, अरब (सौ करोड़ नहीं), साठ अरब (छह हजार करोड़ नहीं) पाँच खरब (पचास हजार करोड़ नहीं) आदि-आदि।

b. Ordinals

Insofar as ordinals are concerned, clear recommendations are given as to which form needs to be used:

पहला	दूसरा	तीसरा	चौथा	पाँचवा
छठा (छठवाँ नहीं)	सातवाँ	आठवाँ	नवाँ	दसवाँ
प्रथम	द्वितीय	तृतीय	चतुर्थ	पंचम
षष्ठ (षष्ठम नहीं)	सप्तम	अष्टम	नवम	दशम

c. Fractions:

Fractions pose a particular problem in Hindi because of the variations in their representation. The following recommendations clearly specify the way Fractional numbers need to be written.

एक चौथाई	$\frac{1}{4}$	सवा दो	$2\frac{1}{4}$
आधा	$\frac{1}{2}$	ढाई (साढ़े दो नहीं)	$2\frac{1}{2}$
पौन	$\frac{3}{4}$	पौने तीन	$2\frac{3}{4}$
सवा (सवा एक नहीं)	$1\frac{1}{4}$	सवा तीन	$3\frac{1}{4}$
डेढ़ (साढ़े एक नहीं)	$1\frac{1}{2}$	साढ़े तीन	$3\frac{1}{2}$ आदि।
पौने दो	$1\frac{3}{4}$		

Similar recommendations are given for fractions used to denote time

पौन बजा है	अर्थात्	12.45
सवा बजा है	अर्थात्	1.15
डेढ़ बजा है	अर्थात्	1.30
पौने दो बजे हैं	अर्थात्	1.45
सवा दो बजे हैं	अर्थात्	2.15
ढाई बजे हैं	अर्थात्	2.30
पौने तीन बजे हैं	अर्थात्	2.45
साढ़े तीन बजे हैं	अर्थात्	3.30

6. RECOMMENDATIONS RE. PARAGRAPH FORMATTING:

With the growth of word-processors and the rising need to write scholarly articles and papers in Hindi, an urgent need is felt to provide norms for formatting of paragraphs. Should Indian numbers and letters be used ? The Central Hindi Directorate provides clear guidelines as to the format of paragraphs and the choice of numbers, as shown below:

I.	वर्णविचार	1.0
क.	स्वर	1.1
1.	स्वर की परिभाषा	1.1.1
2.	स्वर-भेद	1.1.2
1.	रचना के अनुसार	1.1.2.1
(i)	मूल (एकस्वरक)	1.1.2.1.1
(ii)	दीर्घाकृत	1.1.2.1.2
(क)	दीर्घ	1.1.2.1.2.1
(ख)	प्लुत	1.1.2.1.2.2
(iii)	संध्यक्षर	1.1.2.1.3
2.	अनुनासिकता के आधार पर	1.1.2.2
(i)	मौखिक/निरनुनासिक	1.1.2.2.1
(ii)	अनुनासिक	1.1.2.2.2
ख.	व्यंजन	1.2
II	शब्दविचार	2.0
III	वाक्यविचार	3.0
IV	रचना	4.0