

“DEVANĀGARĪ SCRIPT BEHAVIOUR FOR HINDI”

(For ICT stakeholders)

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)

PART A

1. INTRODUCTION	4
2. OBJECTIVES OF "Devanāgarī Script behaviour for Hindi"	7
3. END USERS FOR "Devanāgarī Script behaviour for Hindi"	9
4. SCOPE	9
5. TERMINOLOGY	11
6. PHILOSOPHY AND UNDERLYING PRINCIPLES	15
7. “Devanāgarī Script behaviour for Hindi” : STRUCTURE.....	16
7.1 PERIPHERAL ELEMENTS OF THE "Devanāgarī Script behaviour for Hindi" 17	
7.2 CONFORMITY TO THE SYLLABLE STRUCTURE	19
7.3 “DEVANĀGARĪ SCRIPT BEHAVIOUR FOR HINDI” PROPER	22
7.3.1 The Character Set of Hindi.	22
7.3.2 Consonant Mātrā Combinations.	32
7.3.3 The Ligature Set of Hindi.	48
7.3.4 Valid Combinations And Invalid Combinations.....	61
7.3.5 The Collation Order of Hindi.....	62
8. REFERENCES	64
9. ANNEXURES	65
Annexure 1 : Names of experts who contributed in evolving the draft standard.	65
Annexure 2: Shapes of Hindi characters as per Central Hindi Directorate	67
Annexure 3: Unicode Table of Devanāgarī as per UNICODE 7.0.....	69
Annexure 4: Note On Ambiguous Use Of Anuswara (Anunasika) With Superscript Vowel Signs In Hindi.....	69

PART B

RECOMMENDATIONS FOR DEVELOPERS	71
1. TECHNICAL ISSUES.....	72
1.1. DEVANĀGARĪ SCRIPT AN OVERVIEW	72
1.2. DEVANĀGARĪ: A HISTORICAL PERSPECTIVE	72
1.3. ENCODING PRINCIPLES	73
1.4. THE FUNDAMENTAL UNIT: SYLLABLE	73
1.5. UNICODE TEXT SEGMENTATION UAX #29.....	73
1.6. RENDERING.....	74

1.7.	ZWJ/ZWNJ IN HINDI.....	75
1.8.	CURSOR MOVEMENT& DELETION.....	76
1.9.	NORMALIZATION	78
1.	LINGUISTIC ISSUES.....	80
2.1.	PRELIMINARIES	80
2.2.	CORRECT USAGE	80
a.	Recommendations for affixing the कारक चिह्न:.....	80
b.	Recommendations for Hyphenation Section 3.4. of the document.	80
c.	Recommendations for अनुस्वार and चंद्रबिंदु Section 3.6 of the document.	81
d.	Recommendations regarding the Visarga Section 3.7 of the document.....	83
e.	Recommendations regarding use of the Halant Section 3.8. of the document...	84
f.	Recommendations for use of ऐ & औ cf. Section 3.10 of the document.	85
g.	Miscellaneous Recommendations	85
h.	Recommendations for usage of यव.....	86
3.	LOAN-WORDS AND THEIR CORRECT REPRESENTATION	87
4.	PUNCTUATIONS.....	91
5.	CORRECT SPELLINGS OF NUMERALS:.....	103
6.	RECOMMENDATIONS RE. PARAGRAPH FORMATTING:	105
	REFERENCES	106

PART A

1. INTRODUCTION

The term “**Devanāgarī Script Behaviour for Hindi (for ICT stakeholders)**”, refers to the behaviour pattern of the writing system of Hindi. The term “Devanāgarī Script Behaviour for Hindi (for ICT stakeholders)”, refers to the behaviour pattern of the writing system of Hindi. Languages, which have written representations use a coherent pattern which is similar to the linguistic grammar of a given language of storing the information within the system.

The “Devanāgarī Script behaviour for Hindi” deals with the surface structure of Hindi and tries to provide the best possible “fit” for shapes and their representation of the conjunct forms with the help of specialists (linguists, font designers, language experts, academicians) who work in the area of the written representation of the language. 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. Other languages which use Devanāgarī script, the script grammar for these are dealt with in separate documents.

Devanāgarī 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 Devanāgarī 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 Devanāgarī code-page. Thus Marathi and Konkani use ष and ॠ (for generating out the eyelash ra). These are not present in Hindi or Dogri. The Hindi ॠ (U+090D) is represented in Marathi and Konkani as ञ (U+0972). Nukta is used in Hindi and Dogri but not in Marathi or Konkani.

- b. The shape of the given character.

Although Marathi and Hindi share the same script Devanāgarī, not only do they not share the same character inventory but in addition the representation of certain characters is different. Thus the Hindi /la/ is different from the Marathi /la/ in so far as the placement of the stem is concerned.

Hindi /ल/ Marathi /ल/ .

The same is the case with श्च which is represented in Hindi as श्च but in Marathi as श्च. 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 Marathi as in the following examples¹

Hindi: शक्ति

Hindi: खट्टा

Nepali: शक्ति

Nepali/Marathi: खट्टा

Similarly, in the case of the following stack, a marked difference is noticed. The Central Hindi Directorate states that the conjunct of द and व be linear and not stacked unlike Marathi where the conjunct is stacked:

Hindi: द्वारा

Nepali/Marathi: द्वारा

- 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

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

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

In Nepali क्ष ज्ञ त्र are sorted at the end

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

¹ Although the Central Hindi Directorate recommends क्त as the normative form, in Hindi both versions seem to be used cf. books published by reputed publishers (शक्ति & शक्ति). It would appear that both forms are acceptable in Hindi writing style, but क्त, क्ल (शक्ति, शुक्ल) are standard forms.

² Cf. 6.3.5 of the document

Other languages which use Devanāgarī script, the script grammar for these are dealt with in separate documents.

“Devanāgarī 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 (Hindi, Marathi, Nepali, Sanskrit, Konkani, Maithili, Sindhi, Bodo, Dogri, Santhali, Kashmiri) of India. Each of these languages has different representations of the shape of Individual characters as well as their ligatural representations. The “Devanāgarī Script behaviour for Hindi” in this document is pertinent to Hindi. Other similar documents define the structure of Marathi, Sanskrit, Nepali to name a few languages sharing the common script Devanāgarī and the Unicode code-block 0900-097F (with exception of Rupee Sign and Swastik).

2. OBJECTIVES OF "Devanāgarī Script behaviour for Hindi"

The objectives of the “Devanāgarī 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:

- Abjad, Akshar (Alphasyllabary) Classify the language in terms of its mnemonic as defined in ISO.
- Provide an inventory of the characters pertinent to the language and classify the same in terms of their taxonomy.
- Provide syllabic structure of the language.
- Since Brahmi is written from left to right, and since certain characters do not follow the linear left to right order, 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/Application 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 affect search as well as Natural Language Processing. (refer PART B 1.7 below)

- 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 other languages although languages share the same script. Thus, in Marathi क्ष, ज्ञ are placed at the end of the consonant inventory, i.e. after ह in the sort order. In Hindi क्ष is sorted along with क and ज्ञ with ज.

3. END USERS FOR "Devanāgarī Script behaviour for Hindi"

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

- The primary aim of the “Devanāgarī Script behaviour for Hindi” has been for the font developer. It can be used by font developers desirous of developing a font which is compliant with the perception of the characters and ligatures of a language by its user community.
- 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 “Devanāgarī 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 “Devanāgarī 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 input mechanism using standard keyboard layouts, which will meet the requirement of the particular linguistic community.
- The collation or sort order of Hindi as described in a section 6.3.5 of “Devanāgarī Script behaviour for Hindi” permits the software developer to write software functions/ routines for sorting data in all applications.

Thus this document has a wide range of use and can be of utility to font developers, Indian language developers and linguists in the area of computation.

4. 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 sorting order and its compliance with CLDR
 - Compliance of the script with Unicode

These will be treated within the relevant sections of the document.

5. 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 æ 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 - ऌ [U+ 0933] which Hindi does not use. Flaps used in Hindi ढ [U+095C] ढ [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 ऋ 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.

Dependent Vowel Signs (Matras): The dependent vowels serve as the common manner of writing non-inherent vowels and are generally referred to as vowel signs, or as matras in Sanskrit. The dependent vowels do not stand alone; rather, they are visibly depicted in combination with a base letterform.

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

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

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.

Rendering Order: Devanāgarī and other Indic scripts have some dependent vowels that must be depicted to the left side of their consonant letter, the software that renders the Indic scripts must be able to reorder elements in mapping from the logical (character) store to the presentational (glyph) rendering.

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.

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.

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

Ligature: (see **Conjunct**).

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

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

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

Pure consonant⁹: A consonant which does not have any vowel implicitly associated with it.

Rafar/Repha: 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 = क̣. 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 = छ̣.

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. क̣. 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. क̣̣.

Roman script¹⁰: 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. However Orthographic syllable required for text processing need not correspond exactly with a phonological syllable.

⁹ The term used is as per BIS DOCUMENT *IS 13194: 1991*. However, it could also be termed as “short obstructive sound” as per suggestion of experts.

¹⁰ The term used is as per BIS DOCUMENT *IS 13194: 1991.Point No 2.1.20 p. 1*

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

Vowel: *It 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 syllable¹¹.*

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

¹¹ <http://oxforddictionaries.com/definition/english/vowel>

6. PHILOSOPHY AND UNDERLYING PRINCIPLES

The “Devanāgarī Script behaviour for Hindi” is based on the following principles:

1. The document aims to depict the surface grammar of the written language: the manner in which characters as well as conjuncts are depicted.
2. Where a given script admits many languages, it is pre-supposed that such languages will prescribe different representations for a given shape or conjunct according to the perception of the native users of that language.
3. Corollary to the above, the result is a script and alloscripts i.e. a given script shared by many languages is not uniformly deployed across all the languages, but is subject to variations and modulations.
4. The term “Devanāgarī 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 occur¹².
5. The “Devanāgarī 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.

¹² It is recommended that such variations be culled by placing the document for public review..

7. “Devanāgarī Script behaviour for Hindi” : STRUCTURE

The “Devanāgarī Script behaviour for Hindi” provided below has the following parts.

Part 7.1.

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

Part 7.2.

describes about the syllabic structure.

Part 7.3.

is the “Devanāgarī 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 7.3.1. which deals with the character set of the language. Sections 7.3.2 and 7.3.3 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 Anuswara¹³ and Candrabindu are also provided.

¹³ 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)

7.1 PERIPHERAL ELEMENTS OF THE "Devanāgarī Script behaviour for Hindi"

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. These constitute the elements that are peripheral to the document.

7.1.1. 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*

7.1.2. 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 Indic Orthographic Syllable as its core.

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

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

Abugidas / Alphasyllabaries because of their syllabic structure require a special description which is the subject of the discussion in 7.2. below.

7.1.3. Basic Alphabet System

(C) Consonants

क	क्र	ख	ख	ग
ग	घ	ङ	च	छ
ज	झ	झ	ञ	ट
ठ	ड	ड़	ढ	ढ़
ण	त	थ	द	ध
न	प	फ	फ़	ब
भ	म	य	र	ल

व	श	ष	स	ह
---	---	---	---	---

(V) Vowels^{14 15}

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

(M) Mātrās

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

(D) Diacritics

ं - Anuswara	Anuswara, an archinasal, is denoted by a dot above the letter after which it is to be pronounced. This falls under Nasal category.
ँ - Candrabindu	Candrabindu is pure nasalization as air comes from the nose. It is denoted by a breve with a dot superposed above 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.
ऌ - Avagraha ¹⁶	For extra length with long vowels as seen in the Sanskrit text /उपदेशेऽजनुनासिक/

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

(N)¹⁷ **Nukta** ँ – Nukta is used in Hindi

¹⁴ अं अः 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 अ and ँ ः respectively. It is for this reason that Unicode does not provide separate code points for अं अः.

¹⁵ ऐ औ are accepted for transliterating Loan words borrowed from English such as ऑफ. In the case of the two short vowels ऐ, औ used in Dravidian, as per recommendation of Central Hindi Directorate 3.15.6.1.2 p. 35, words having a short letter/short vowel sign and admitted in Hindi are automatically replaced by the corresponding long letter/vowel sign. Thus the short “o” of Oddanchatram ஒட்டன்சத்திரம் is rendered as long o in Hindi: ओट्टणचत्तिरम

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

¹⁷ The nukta is a small dot placed under a character in certain scripts to show that they are flapped or for deriving consonants required for Urdu क, ख, ग, ज, फ

7.2 CONFORMITY TO THE SYLLABLE STRUCTURE

The syllable is a basic unit of speech studied on both the phonetic and phonological levels of analysis. It is very easy for a native language speaker to count the number of syllables in a sequence however Orthographic syllable required for text processing need not correspond exactly with a phonological syllable.

ABNF based definition of Indic Orthographic Syllable

Augmented Backus–Naur Form (ABNF) is a meta-language based on Backus–Naur Form (BNF), but consisting of its own syntax and derivation rules. The motive principle for ABNF is to describe a formal system of a language to be used as a bidirectional communications protocol.

$V[m] \mid \{CH\}C[v][m]CH$

The linguistic definition of Indic syllable has been mapped to ABNF(Augmented Backus–Naur Form) for the purpose of text segmentation, line breaking , drop letter, letter spacing in horizontal text and vertical text representation. The definition has been elaborated , taking Hindi as an example.

The definition is a combination of 3 rules :

Rule 1 : $V[m]$

Rule 2 : $\{CH\}C[v][m]$

Rule 3 : CH (This rule is applicable only at the end of the word)

V (upper case) is independent vowel

m is modifier(Anusvara/Visarga/Chandrabindu)

C is a consonant as per Unicode's definition which may or may not include a single nukta

v (lower case) is any dependent vowel or vowel sign (mātrā)

H is halant / virama (as per UNICODE)

$|$ is a rule separator

$[]$ - The enclosed items is optional under this bracket

$\{ \}$ - The enclosed item/items occurs zero or repeated multiple times

Various Use cases of ABNF based Indic Syllable definition for Hindi language as example

Rule 1 : V[m]

Sl. No.	Examples	Definition
1	अ, ई, उ	V (Vowel) is a syllable
2	अं, उँ, आः	V+ Modifier is a syllable

Rule 2 : {CH}C[v][m]

Sl. No.	Examples	Definition
1	र, क, ज, ल, म	Consonant is a syllable
2	प्प, क्ख, च्च, ज्ज्व, त्क्ल, त्स्न	Zero or more Consonant + Virama sequences followed by consonant is a syllable
3	र्त, त्स, त्स्न, त्स्न्य, फ़क्क	Zero or more Consonant (Nukta) +Virama followed by consonant is a syllable
4	र्ता, त्स्न्या, फ़जी, क्या	Zero or more consonant+ (Nukta)+ virāma sequences followed by a consonant (+Nukta) followed by a vowel sign is a syllable
5	तः, स्तः, स्त्रँ, स्तः, फ़ज़ँ	zero or more consonant+ (Nukta)+ virāma sequences followed by a consonant (+Nukta) followed by modifier is a syllable
6	त्स्न्याः त्स्न्युं, त्स्न्युं, फ़ज़ें, हिं	zero or more consonant+ (Nukta)+ virāma sequences followed by a consonant (+Nukta) followed by a vowel sign and modifier is a syllable
7	स्थि, जि, ख्वा	Zero or more Consonant +halant sequences followed by a consonant followed by vowel sign is a syllable

Rule 3 : CH

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

Examples of combination of the rules :

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

स्वा	$CHCv$
ग	C
त	C
म्	CH

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

भ	C
र	C
त	C
ना	Cv
ट्य	CHC
म	C

7.3 “DEVANĀGARĪ SCRIPT BEHAVIOUR FOR HINDI” PROPER

This section lays down in detail the different parameters of the “Devanāgarī Script Behaviour for Hindi”. These are:

7.3.1. The Character Set of Hindi.

7.3.2. The Consonant mātrā combinations of Hindi as well as Vowel and nasal modifier combinations.

7.3.3. The Ligature Set of Hindi.

7.3.4 Inventory of Valid and Invalid Combinations with respect to 7.3.2. and 7.3.3.

7.3.5. Collocation Order of Hindi

7.3.1 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 “Devanāgarī Script behaviour for Hindi”.

This comprises the following:

7.3.1.1. The Consonant Set.¹⁸

7.3.1.2. The Vowel Set.

7.3.1.3. The Mātrā Set.

7.3.1.4. Memory Representation and Rendering Order.

7.3.1.5. Shape of the combination of ra (rakar, repha).

7.3.1.6. The Set of Diacritics.

7.3.1.7. Halant

7.3.1.8. Numerals.

7.3.1.9. Punctuation marks.

7.3.1.10. Other symbols.

¹⁸ 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:

7.3.1.1. The Consonant Set

The Consonant set of Hindi comprises the following characters:

A basic Consonant inventory arranged as per their Vargas.

Character	Unicode code-point	Character name
क	U+0915	DEVANAGARI LETTER KA
ख	U+0916	DEVANAGARI LETTER KHA
ग	U+0917	DEVANAGARI LETTER GA
घ	U+0918	DEVANAGARI LETTER GHA
ङ	U+0919	DEVANAGARI LETTER NGA
च	U+091A	DEVANAGARI LETTER CA
छ	U+091B	DEVANAGARI LETTER CHA
ज	U+091C	DEVANAGARI LETTER JA
झ	U+091D	DEVANAGARI LETTER JHA
ञ	U+091E	DEVANAGARI LETTER NYA
ट	U+091F	DEVANAGARI LETTER TTA
ठ	U+0920	DEVANAGARI LETTER TTHA
ड	U+0921	DEVANAGARI LETTER DDA
ढ	U+0922	DEVANAGARI LETTER DDHA
ण	U+0923	DEVANAGARI LETTER NNA
त	U+0924	DEVANAGARI LETTER TA
थ	U+0925	DEVANAGARI LETTER THA
द	U+0926	DEVANAGARI LETTER DA
ध	U+0927	DEVANAGARI LETTER DHA
न	U+0928	DEVANAGARI LETTER NA
न	U+0929	DEVANAGARI LETTER NNA
प	U+092A	DEVANAGARI LETTER PA
फ	U+092B	DEVANAGARI LETTER PHA
ब	U+092C	DEVANAGARI LETTER BA
भ	U+092D	DEVANAGARI LETTER BHA
म	U+092E	DEVANAGARI LETTER MA
य	U+092F	DEVANAGARI LETTER YA

र	U+0930	DEVANAGARI LETTER RA
ॠ	U+0931	DEVANAGARI LETTER RRA
ल	U+0932	DEVANAGARI LETTER LA
ळ	U+0933	DEVANAGARI LETTER LLA
ॡ	U+0934	DEVANAGARI LETTER LLLA
व	U+0935	DEVANAGARI LETTER VA
श	U+0936	DEVANAGARI LETTER SHA
ष	U+0937	DEVANAGARI LETTER SSA
स	U+0938	DEVANAGARI LETTER SA
ह	U+0939	DEVANAGARI LETTER HA
क्र	U+0958	DEVANAGARI LETTER QA
ख्र	U+0959	DEVANAGARI LETTER KHHA
ग़	U+095A	DEVANAGARI LETTER GHHA
ज़	U+095B	DEVANAGARI LETTER ZA
ड़	U+095C	DEVANAGARI LETTER DDDHA
ढ़	U+095D	DEVANAGARI LETTER RHA
फ़	U+095E	DEVANAGARI LETTER FA
य़	U+095F	DEVANAGARI LETTER YYA

Note: Ligatures क्षत्रज्ञश्र are not listed in the consonants list, since they are ligatures: क्ष= क+्+ष, ज्ञ=ज+्+ञ, त्र=त+्+र, श्र=श+्+र. 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.

7.3.1.2. The Vowel Set

The Vowel set of Hindi is as under:

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

7.3.1.3. The Mātrā Set

The Mātrā¹⁹ (Vowel Sign) of Hindi is as under:

Mātrā Names	Mātrās Sign	Where is it used ?	Consonant Shapes formed
DEVANĀGARĪ VOWEL SIGN AA	ा	आ	क् + आ = का
DEVANĀGARĪ VOWEL SIGN I	ि	इ	क् + इ = कि
DEVANĀGARĪ VOWEL SIGN II	ी	ई	क् + ई = की
DEVANĀGARĪ VOWEL SIGN U	ु	उ	क् + उ = कु
DEVANĀGARĪ VOWEL SIGN UU	ू	ऊ	क् + ऊ = कू
DEVANĀGARĪ VOWEL SIGN VOCALIC R	ृ	ऋ	क् + ऋ = कृ
DEVANĀGARĪ VOWEL SIGN E	े	ए	क् + ए = के
DEVANĀGARĪ VOWEL SIGN CANDRA E	ँ	ऐ	क् + ऐ = कै
DEVANĀGARĪ VOWEL SIGN AI	ै	ऐ	क् + ऐ = कै
DEVANĀGARĪ VOWEL SIGN O	ो	ओ	क् + ओ = को
DEVANĀGARĪ VOWEL SIGN CANDRA O	ॉ	ऑ	क् + ऑ = कौ
DEVANĀGARĪ VOWEL SIGN AU	ौ	औ	क् + औ = कौ

¹⁹ Allographs

7.3.1.4. Memory Representation and Rendering Order

The storage of plain text in Devanāgarī and all other Indic scripts generally follows phonetic order; that is, a consonant with vowel sign is a syllable with a dependent vowel is always encoded as a consonant letter followed by a vowel sign in the memory representation. (see below Figure).

Figure. Rendering Order in Devanāgarī

Character Order Glyph Order

क + ि → कि

Because Devanāgarī and other Indic scripts have some dependent vowels that must be depicted to the left side of their consonant letter, the software that renders the Indic scripts must be able to reorder elements in mapping from the logical (character) store to the presentational (glyph) rendering. For example, if C denotes the nominal form of consonant, and v denotes a left-side dependent vowel sign form of vowel V, then a reordering of glyphs with respect to encoded characters occurs as just shown.

When the dependent vowel is used to override the inherent vowel of a syllable, it is always written to the extreme left of the orthographic syllable. If the orthographic syllable contains a consonant cluster, then this vowel is always depicted to the left of that cluster.

त् + र + ि → त्र + ि → त्रि

7.3.1.5. Shape of the combination of ra (Rakar, Repha/ Rafar)

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:

REPHA/RAFARS

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

धर्म, चर्खा

RAKARS

1. Bottom rakar ढ़ ड़
2. Side rakar क्र च्र
3. Inside rakar ह्र

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

Bottom rakar	ड़म, राष़्ट्र
Side rakar	व्रत, चक्र, प्रसाद, स्रोत
Inside rakar	ह्रस्व, हास

7.3.1.6. Diacritics

These are as under in the case of Hindi:

- ं - Anuswar रंग
- ँ - Candrabindu/Anunasika ईँट
- ः - Visarga दुःख
- ॆ - For extra length with long vowels e.g. /उपदेशेऽजनुनासिक / ²⁰

7.3.1.7. Halant

- ् - Halant संवत्

7.3.1.8. Numerals

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

० १ २ ३ ४ ५ ६ ७ ८ ९

Numeral Shapes	Explanation
०	Devanāgarī Digit Zero

²⁰ 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>

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

7.3.1.9. 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 DEVANĀGARĪ DANDA (।) and DEVANĀGARĪ DOUBLE DANDA (॥).

। is used to mark the full stop and । and ॥ are used for poetry written in छंद /दोहा / चौपाई (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 रौंदूंगी तोय ॥

A list of punctuations is provided below:

Devanāgarī Name	Name of the marker	Marker Shape
"स्टॉप" ²¹ /stop/	U+002E FULL STOP	.
प्रश्न चिह्न /prashna chihna/	U+003F QUESTION MARK	?
अल्पविराम /alpaviram/	U+002C COMMA	,
विस्मयसूचक चिह्न /vismaysuchak chihna/	U+0021 EXCLAMATION MARK	!
ऊर्ध्व अल्पविराम /urdhva alpaviram/	U+0027 APOSTROPHE	'
अर्धविराम /ardhaviram/	U+003B SEMICOLON	;
उपविराम /upviram/	U+003A COLON	:
योजक चिह्न /yojak chihna/	U+002D HYPHEN-MINUS	-
निर्देशक चिह्न /nirdeshak chihna/	U+002D HYPHEN-MINUS U+002D HYPHEN-MINUS Dash	--
लोप चिह्न	FULL STOP (U+002E) FULL STOP (U+002E) FULL STOP (U+002E)	...

²¹ [http://hi.wikipedia.org/wiki/विराम_\(चिन्ह\)](http://hi.wikipedia.org/wiki/विराम_(चिन्ह))

/lop chihna/	Elimination Sign	
	U+002F SLASH	/
उद्धरण चिह्न /udharan chihna/	U+0022 QUOTATION MARK	" "
शब्द चिह्न /shabda chihna	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+002D QUOTATION MARK U+002D HYPHEN-MINUS U+002D HYPHEN- MINUS as above	-- " --
कोष्ठक /koshtak/	U+0028 LEFT PARENTHESIS U+0029 RIGHT PARENTHESIS	()
कोष्ठक /koshtak/	U+005B LEFT SQUARE BRACKET U+005D RIGHT SQUARE BRACKET	[]
कोष्ठक /koshtak/	U+007B LEFT CURLY BRACKET U+007D RIGHT CURLY BRACKET	{ }
संक्षेपसूचक चिह्न /sankshepsuchak chinha/	U+0970 DEVANĀGARĪ ABBREVIATION SIGN	◌
पूर्ण विराम /purna viram/	U+0964 DEVANĀGARĪ DANDA Devanāgarī Purna Viram	
दीर्घ विराम /dirgha viram/	U+0965 DEVANĀGARĪ DOUBLE DANDA Devanāgarī Deergha Viram	॥
हंसपद / त्रुटिबोधक /hanspad trutibodhak/	Sign of left/Add word	↵

7.3.1.10 Other Symbols

These are religious symbols and currency symbol included in Unicode:

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

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

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

7.3.2 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 as shown in set (Table).

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. Therefore, the “Devanāgarī Script Behaviour for Hindi” document provides such combinations which could be termed as “dead” but which are required by the font developer for developing the font for the particular language.

Due to constraints of space and also for reasons of clarity, for each class a series of 4 sets are provided.

Set 1:	क	ख	ग	घ	ङ	च	छ	ज	झ	ञ
Set 2:	ट	ठ	ड	ढ	ण	त	थ	द	ध	न
Set 3:	प	फ	ब	भ	म	य	र	ल	व	श
	ष	स	ह							
Set 4:	क्र	ख्र	ग़	ज़	फ़	ड़	ढ़			

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

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

The classes are as under:

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

7.3.2.2. 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.

7.3.2.1 Consonant and Mātrā combinations.

This set refers to a simple concatenation of consonant and mātrā.

Consonant and Mātrā combinations Set 1

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

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

Consonant and Mātrā combinations Set 2

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

	ट	ठ	ड	ढ	ण	त	थ	द	ध	न
ा	टा	ठा	डा	ढा	णा	ता	था	दा	धा	ना
ि	टि	ठि	डि	ढि	णि	ति	थि	दि	धि	नि
ी	टी	ठी	डी	ढी	णी	ती	थी	दी	धी	नी
ु	टु	ठु	डु	ढु	णु	तु	थु	दु	धु	नु
ू	टू	ठू	डू	ढू	णू	तू	थू	दू	धू	नू
ृ	टृ	ठृ	डृ	ढृ	णृ	तृ	थृ	दृ	धृ	नृ
े	टे	ठे	डे	ढे	णे	ते	थे	दे	धे	ने
ै	टै	ठै	डै	ढै	णै	तै	थै	दै	धै	नै
ो	टो	ठो	डो	ढो	णो	तो	थो	दो	धो	नो
ॉ	टॉ	ठॉ	डॉ	ढॉ	णॉ	तॉ	थॉ	दॉ	धॉ	नॉ
ौ	टौ	ठौ	डौ	ढौ	णौ	तौ	थौ	दौ	धौ	नौ

Consonant and Mātrā combinations Set 3

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

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

Consonant and Mātrā combinations Set 4

This set is in continuation of set 3 which shows consonant and mātrā combinations for nukta consonants.²²

	क़	ख़	ग़	ज़	फ़	ड़	ढ़
ा	का	खा	गा	जा	फा	डा	ढा
ि	कि	खि	गि	जि	फि	डि	ढि
ी	की	खी	गी	जी	फी	डी	ढी
ु	कु	खु	गु	जु	फु	डु	ढु
ू	कू	खू	गू	जू	फू	डू	ढू
ृ ²³	कृ	खृ	गृ	जृ	फृ	डृ	ढृ
े	के	खे	गे	जे	फे	डे	ढे
ै	कै	खै	गै	जै	फै	डै	ढै
ो	को	खो	गो	जो	फो	डो	ढो
ॉ	कॉ	खॉ	गॉ	जॉ	फॉ	डॉ	ढॉ
ौ	कौ	खौ	गौ	जौ	फौ	डौ	ढौ

²² क़, ग़ were not recognized as part of the consonant set by the Central Hindi Directorate, but have been accepted in its latest version..

²³ This combination is rarely used but still noted as a guide for the font designer.

7.3.2.2 Consonant and Mātrā +Nasal combinations.

This set refers to a consonant and mātrā + nasal marker combinations.

Consonant and Mātrā + Nasal combinations - Set 1

	क	ख	ग	घ	ङ	च	छ	ज	झ	ञ
ं	कं	खं	गं	घं	ङं	चं	छं	जं	झं	ञं
ां	कां	खां	गां	घां	ङां	चां	छां	जां	झां	जां
िं	किं	खिं	गिं	घिं	ङिं	चिं	छिं	जिं	झिं	ञिं
ीं	कीं	खीं	गीं	घीं	ङीं	चीं	छीं	जीं	झीं	जीं
ुं	कुं	खुं	गुं	घुं	ङुं	चुं	छुं	जुं	झुं	जुं
ूं	कूं	खूं	गूं	घूं	ङूं	चूं	छूं	जूं	झूं	जूं
ृं	कृं	खृं	गृं	घृं	ङृं	चृं	छृं	जृं	झृं	जृं
ें	कें	खें	गें	घें	ङें	चें	छें	जें	झें	जें
ैं	कैं	खैं	गैं	घैं	ङैं	चैं	छैं	जैं	झैं	जैं
ौं	कौं	खौं	गौं	घौं	ङौं	चौं	छौं	जौं	झौं	जौं
ॉं	काँ	खाँ	गाँ	घाँ	ङाँ	चाँ	छाँ	जाँ	झाँ	जाँ
ौं	काँ	खाँ	गाँ	घाँ	ङाँ	चाँ	छाँ	जाँ	झाँ	जाँ

Consonant and Mātrā +Nasal combinations - Set 2

This set is in continuation of set 1 above which shows combinations of consonant and mātrā + nasal marker

	ट	ठ	ड	ढ	ण	त	थ	द	ध	न
ं	टं	ठं	डं	ढं	णं	तं	थं	दं	धं	नं
ां	टां	ठां	डां	ढां	णां	तां	थां	दां	धां	नां
िं	टिं	ठिं	डिं	ढिं	णिं	तिं	थिं	दिं	धिं	निं
ीं	टीं	ठीं	डीं	ढीं	णीं	तीं	थीं	दीं	धीं	नीं
ुं	टुं	ठुं	डुं	ढुं	णुं	तुं	थुं	दुं	धुं	नुं
ूं	टूँ	ठूँ	डूँ	ढूँ	णूँ	तूँ	थूँ	दूँ	धूँ	नूँ
ृं	ट्रं	ठ्रं	ड्रं	ढ्रं	ण्रं	त्रं	थ्रं	द्रं	ध्रं	न्रं
ेँ	टैँ	ठैँ	डैँ	ढैँ	णैँ	तैँ	थैँ	दैँ	धैँ	नैँ
ँ ²⁴	टँ	ठँ	डँ	ढँ	णँ	तँ	थँ	दँ	धँ	नँ
ैँ	टैं	ठैं	डैं	ढैं	णैं	तैं	थैं	दैँ	धैं	नैं
ोँ	टों	ठों	डों	ढों	णों	तों	थों	दों	धों	नों
ॉँ	टॉँ	ठॉँ	डॉँ	ढॉँ	णॉँ	तॉँ	थॉँ	दॉँ	धॉँ	नॉँ
ौँ	टौँ	ठौँ	डौँ	ढौँ	णौँ	तौँ	थौँ	दौँ	धौँ	नौँ

²⁴ ँ,ँ are used only for the purpose of writing loan words mainly from English. However, since such occurrences are rare, the nasal shapes of ँ,ँ are rarely encountered in Hindi.

Consonant and Mātrā +Nasal combinations - Set 3

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

	प	फ	ब	भ	म	य	र	ल	व	श	ष	स	ह
ं	पं	फं	बं	भं	मं	यं	रं	लं	वं	शं	षं	सं	हं
ां	पां	फां	बां	भां	मां	यां	रां	लां	वां	शां	षां	सां	हां
िं	पिं	फिं	बिं	भिं	मिं	यिं	रिं	लिं	विं	शिं	षिं	सिं	हिं
ीं	पीं	फीं	बीं	भीं	मीं	यीं	रीं	लीं	वीं	शीं	षीं	सीं	हीं
ुं	पुं	फुं	बुं	भुं	मुं	युं	रुं	लुं	वुं	शुं	षुं	सुं	हुं
ूँ	पूँ	फूँ	बूँ	भूँ	मूँ	यूँ	रूँ	लूँ	वूँ	शूँ	षूँ	सूँ	हूँ
ृं	पृं	फृं	बृं	भृं	मृं	यृं	रृं	लृं	वृं	शृं	षृं	सृं	हृं
ेँ	पेँ	फेँ	बेँ	भेँ	मेँ	येँ	रेँ	लेँ	वेँ	शेँ	षेँ	सेँ	हेँ
ँ ²⁵	पँ	फँ	बँ	भँ	मँ	यँ	रँ	लँ	वँ	शँ	षँ	सँ	हँ
ैँ	पैँ	फैँ	बैँ	भैँ	मैँ	यैँ	रैँ	लैँ	वैँ	शैँ	षैँ	सैँ	हैँ
ोँ	पोँ	फोँ	बोँ	भोँ	मोँ	योँ	रोँ	लोँ	वोँ	शोँ	षोँ	सोँ	होँ
ौँ	पौँ	फौँ	बौँ	भौँ	मौँ	यौँ	रौँ	लौँ	वौँ	शौँ	षौँ	सौँ	हौँ

²⁵ Since ँ,ँ are used only for the purpose of writing loan words mainly from English. However, since such occurrences are rare, the nasal shapes of ँँ,ँँ are rarely encountered in Hindi.

Consonant and Mātrā +Nasal combinations - Set 4

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

	क	ख	ग	ङ	फ	ड	ढ
ं	कं	खं	गं	ङं	फं	डं	ढं
ा	कां	खां	गां	गां	फां	डां	ढां
िं	किं	खिं	गिं	ङिं	फिं	डिं	ढिं
ीं	कीं	खीं	गीं	ङीं	फीं	डीं	ढीं
ुं	कुं	खुं	गुं	ङुं	फुं	डुं	ढुं
ूँ	कूँ	खूँ	गूँ	ङूँ	फूँ	डूँ	ढूँ
ृं	कृं	खृं	गृं	ङृं	फृं	डृं	ढृं
ेँ	केँ	खेँ	गेँ	ङेँ	फेँ	डेँ	ढेँ
ँ ²⁶	कँ	खँ	गँ	ङँ	फँ	डँ	ढँ
ैँ	कैँ	खैँ	गैँ	ङैँ	फैँ	डैँ	ढैँ
ोँ	कोँ	खोँ	गोँ	ङोँ	फोँ	डोँ	ढोँ
ौँ	कौँ	खौँ	गौँ	ङौँ	फौँ	डौँ	ढौँ

²⁶ Since ँ, ौँ are used only for the purpose of writing loan words mainly from English. However, since such occurrences are rare, the nasal shapes of ँ, ौँ are rarely encountered in Hindi.

Consonant and Mātrā + Nasal combinations: With Candrabindu - Set 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 Candrabindu 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
ँ	कँ	खँ	गँ	घँ	ङँ	चँ	छँ	जँ	झँ	ञँ
ाँ	काँ	खाँ	गाँ	घाँ	ङाँ	चाँ	छाँ	जाँ	झाँ	जाँ
िँ	किँ	खिँ	गिँ	घिँ	ङिँ	चिँ	छिँ	जिँ	झिँ	ञिँ
ीँ	कीँ	खीँ	गीँ	घीँ	ङीँ	चीँ	छीँ	जीँ	झीँ	जीँ
ुँ	कुँ	खुँ	गुँ	घुँ	ङुँ	चुँ	छुँ	जुँ	झुँ	जुँ
ूँ	कूँ	खूँ	गूँ	घूँ	ङूँ	चूँ	छूँ	जूँ	झूँ	जूँ
ृँ	कृँ	खृँ	गृँ	घृँ	ङृँ	चृँ	छृँ	जृँ	झृँ	जृँ
ेँ	केँ	खेँ	गेँ	घेँ	ङेँ	चेँ	छेँ	जेँ	झेँ	जेँ
ैँ	कैँ	खैँ	गैँ	घैँ	ङैँ	चैँ	छैँ	जैँ	झैँ	जैँ
ोँ	कोँ	खोँ	गोँ	घोँ	ङोँ	चोँ	छोँ	जोँ	झोँ	जोँ
ाँ	काँ	खाँ	गाँ	घाँ	ङाँ	चाँ	छाँ	जाँ	झाँ	जाँ
ौँ	कौँ	खौँ	गौँ	घौँ	ङौँ	चौँ	छौँ	जौँ	झौँ	जौँ

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.

	ट	ठ	ड	ढ	ण	त	थ	द	ध	न
ँ	टँ	ठँ	डँ	ढँ	णँ	तँ	थँ	दँ	धँ	नँ
ाँ	टाँ	ठाँ	डाँ	ढाँ	णाँ	ताँ	थाँ	दाँ	धाँ	नाँ
िँ	टिँ	ठिँ	डिँ	ढिँ	णिँ	तिँ	थिँ	दिँ	धिँ	निँ
ीँ	टीँ	ठीँ	डीँ	ढीँ	णीँ	तीँ	थीँ	दीँ	धीँ	नीँ
ुँ	टुँ	ठुँ	डुँ	ढुँ	णुँ	तुँ	थुँ	दुँ	धुँ	नुँ
ूँ	टूँ	ठूँ	डूँ	ढूँ	णूँ	तूँ	थूँ	दूँ	धूँ	नूँ
ृँ	ट्रँ	ठ्रँ	ड्रँ	ढ्रँ	ण्रँ	त्रँ	थ्रँ	द्रँ	ध्रँ	न्रँ
ेँ	टेँ	ठेँ	डेँ	ढेँ	णेँ	तेँ	थेँ	देँ	धेँ	नेँ
ैँ	टैँ	ठैँ	डैँ	ढैँ	णैँ	तैँ	थैँ	दैँ	धैँ	नैँ
ोँ	टोँ	ठोँ	डोँ	ढोँ	णोँ	तोँ	थोँ	दोँ	धोँ	नोँ
ॉँ ²⁷	टाँ	ठाँ	डाँ	ढाँ	णाँ	ताँ	थाँ	दाँ	धाँ	नाँ
ौँ	टौँ	ठौँ	डौँ	ढौँ	णौँ	तौँ	थौँ	दौँ	धौँ	नौँ

²⁷ Since ॉँ, ौँ are used only for the purpose of writing loan words mainly from English. However, since such occurrences are rare, the nasal shapes of ॉँ, ौँ are rarely encountered in Hindi.

Consonant and Mātrā +Nasal combinations With Candrabindu - Set 3

This set is in continuation of set 2 above which shows combinations of Consonant and Mātrā + Candrabindu. 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
िँ	पिँ	फिँ	बिँ	भिँ	मिँ	यिँ	रिँ	लिँ	विँ	शिँ	षिँ	सिँ	हिँ
ीँ	पीँ	फीँ	बीँ	भीँ	मीँ	यीँ	रीँ	लीँ	वीँ	शीँ	षीँ	सीँ	हीँ
ुँ	पुँ	फुँ	बुँ	भुँ	मुँ	युँ	रुँ	लुँ	वुँ	शुँ	षुँ	सुँ	हुँ
ूँ	पूँ	फूँ	बूँ	भूँ	मूँ	यूँ	रूँ	लूँ	वूँ	शूँ	षूँ	सूँ	हूँ
ृँ	पृँ	फृँ	बृँ	भृँ	मृँ	यृँ	रृँ	लृँ	वृँ	शृँ	षृँ	सृँ	हृँ
ेँ	पेँ	फेँ	बेँ	भेँ	मेँ	येँ	रेँ	लेँ	वेँ	शेँ	षेँ	सेँ	हेँ
ैँ	पेँ	फेँ	बेँ	भेँ	मेँ	येँ	रेँ	लेँ	वेँ	शेँ	षेँ	सेँ	हेँ
ोँ	पोँ	फोँ	बोँ	भोँ	मोँ	योँ	रोँ	लोँ	वोँ	शोँ	षोँ	सोँ	होँ
ॉँ	पाँ	फाँ	बाँ	भाँ	माँ	याँ	राँ	लाँ	वाँ	शाँ	षाँ	साँ	हाँ
ौँ	पोँ	फौँ	बौँ	भौँ	मौँ	योँ	रौँ	लौँ	वौँ	शौँ	षौँ	सौँ	हौँ

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.

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

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

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

Although क्ष, ज्ञ, त्र and श्र are not consonants per se²⁸, 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
ँ	क्षँ	ज्ञँ	त्रँ	श्रँ
ै	क्षै	ज्ञै	त्रै	श्रै
ो	क्षो	ज्ञो	त्रो	श्रो
ॉ	क्षॉ	ज्ञॉ	त्रॉ	श्रॉ
ौ	क्षौ	ज्ञौ	त्रौ	श्रौ

²⁸ Cf. p. 22

Set 2 Consonant and Mātrā + Nasal combinations: with Anuswara

	क्ष	ज	त्र	श्र
ं	क्षं	जं	त्रं	श्रं
ां	क्षां	जां	त्रां	श्रां
िं	क्षिं	जिं	त्रिं	श्रिं
ीं	क्षीं	जीं	त्रीं	श्रीं
ुं	क्षुं	जुं	त्रुं	श्रुं
ूं	क्षूं	जूं	त्रूं	श्रूं
ृं	क्षृं	जृं	त्रृं	श्रृं
ें	क्षे	जे	त्रे	श्रे
ँ	क्षँ	जँ	त्रँ	श्रँ
ैं	क्षै	जै	त्रै	श्रै
ो	क्षो	जो	त्रो	श्रो
ॉ	क्षॉ	जॉ	त्रॉ	श्रॉ
ौ	क्षौ	जौ	त्रौ	श्रौ

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	ज	त्र	श्र
ँ	क्षँ	जँ	त्रँ	श्रँ
ाँ	क्षाँ	जाँ	त्राँ	श्राँ
िँ	क्षिँ	जिँ	त्रिँ	श्रिँ
ीँ	क्षीँ	जीँ	त्रीँ	श्रीँ
ुँ	क्षुँ	जुँ	त्रुँ	श्रुँ
ूँ	क्षूं	जूँ	त्रूं	श्रूं

ॠँ	ॡँ	ॢँ	ॣँ	।ँ
ेँ	॥ँ	०ँ	ॡँ	ॢँ
ँँ	॥ँ	०ँ	ॡँ	ॢँ
ैँ	॥ँ	०ँ	ॡँ	ॢँ
ोँ	॥ँ	०ँ	ॡँ	ॢँ
ॉँ	॥ँ	०ँ	ॡँ	ॢँ
ौँ	॥ँ	०ँ	ॡँ	ॢँ

7.3.3 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

7.3.3.1. 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 “क्क²⁹”.

CHC(combination of two consonants) - Set 1										
	क	ख	ग	घ	ङ	च	छ	ज	झ	ञ
क्	क्क ³⁰	क्ख	क्ग	क्घ	क्ङ	क्च	क्छ	क्ज	क्झ	क्ञ
ख्	क्ख	क्ख्ख	क्खग	क्खघ	क्खङ	क्खच	क्खछ	क्खज	क्खझ	क्खञ
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	च्च	च्च	च्च
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

²⁹ ZWJ/ZWNJ are generally not used in Hindi. cf. PART B 1.7 : Note on ZWJ/ZWNJ in Hindi.

³⁰ Although the Central Hindi Directorate prefers the linear form of the conjunct, the use of क्क i.e. a stacked form is also permissible and is therefore noted as such.

ट्	ट्क	ट्र	टग	टघ	टङ	टच	टछ	टज	टझ	टञ
ड्	ड्क	ड्र	डग	डघ	डङ	डच	डछ	डज	डझ	डञ
ढ्	ढ्क	ढ्र	ढग	ढघ	ढङ	ढच	ढछ	ढज	ढझ	ढञ
ण्	ण्क	ण्र	णग	णघ	णङ	णच	णछ	णज	णझ	णञ
त्	त्क	त्र	तग	तघ	तङ	तच	तछ	तज	तझ	तञ
थ्	थ्क	थ्र	था	थघ	थङ	थच	थछ	थज	थझ	थञ
द्	द्क	द्र	द्ग	द्घ	द्ङ	द्च	द्छ	द्ज	द्झ	द्ञ
ध्	ध्क	ध्र	धा	धघ	धङ	धच	धछ	धज	धझ	धञ
न्	न्क	न्र	ना	नघ	नङ	नच	नछ	नज	नझ	नञ
प्	प्क	प्र	पा	पघ	पङ	पच	पछ	पज	पझ	पञ
फ्	फ्क	फ्र	फा	फघ	फङ	फच	फछ	फज	फझ	फञ
ब्	ब्क	ब्र	बा	बघ	बङ	बच	बछ	बज	बझ	बञ
भ्	भ्क	भ्र	भा	भघ	भङ	भच	भछ	भज	भझ	भञ
म्	म्क	म्र	मा	मघ	मङ	मच	मछ	मज	मझ	मञ
य्	य्क	य्र	या	यघ	यङ	यच	यछ	यज	यझ	यञ
र्	र्क	र्ष	र्ग	र्घ	र्ङ	र्च	र्छ	र्ज	र्झ	र्ञ
ल्	ल्क	ल्र	ला	लघ	लङ	लच	लछ	लज	लझ	लञ
व्	व्क	व्र	वा	वघ	वङ	वच	वछ	वज	वझ	वञ
श्	श्क	श्र	शग	शघ	शङ	श्च	श्छ	श्ज	श्झ	श्ञ
ष्	ष्क	ष्र	षा	षघ	षङ	ष्च	ष्छ	ष्ज	ष्झ	ष्ञ
स्	स्क	स्र	सा	सघ	सङ	स्च	स्छ	स्ज	स्झ	स्ञ
ह्	ह्क	ह्र	हग	हघ	हङ	हच	हछ	हज	हझ	हञ
क्	क्क	क्व	क्ग	क्घ	क्ङ	क्च	क्छ	क्ज	क्झ	क्ञ
ख्	ख्क	ख्र	खा	खघ	खङ	खच	खछ	खज	खझ	खञ
ग्	ग्क	ग्र	गग	गघ	गङ	ग्च	गछ	गज	गझ	गञ
ज्	ज्क	ज्र	जा	जघ	जङ	ज्च	ज्छ	ज्ज	ज्झ	ज्ञ
फ़्	फ़्क	फ़्र	फ़ा	फ़घ	फ़ङ	फ़च	फ़छ	फ़ज	फ़झ	फ़ञ

इ	इक	इख	इग	इघ	इङ	इच	इछ	इज	इझ	इञ
ढ़	ढ़क	ढ़ख	ढ़ग	ढ़घ	ढ़ङ	ढ़च	ढ़छ	ढ़ज	ढ़झ	ढ़ञ

CHC Set 2:

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

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

ब्र	बट	बठ	बड	बढ	बण	बत	बथ	बद	बध	बन
भ्र	भट	भठ	भड	भढ	भण	भत	भथ	भद	भध	भन
म्र	मट	मठ	मड	मढ	मण	मत	मथ	मद	मध	मन
य्र	यट	यठ	यड	यढ	यण	यत	यथ	यद	यध	यन
र्र	रट	रठ	रड	रढ	रण	रत	रथ	रद	रध	रन
ल्र	लट	लठ	लड	लढ	लण	लत	लथ	लद	लध	लन
व्र	वट	वठ	वड	वढ	वण	वत	वथ	वद	वध	वन
श्र	शट	शठ	शड	शढ	शण	शत	शथ	शद	शध	शन
ष्र	षट	षठ	षड	षढ	षण	षत	षथ	षद	षध	षन
स्र	सट	सठ	सड	सढ	सण	सत	सथ	सद	सध	सन
ह्र	हट	हठ	हड	हढ	हण	हत	हथ	हद	हध	हन
क्र	कट	कठ	कड	कढ	कण	कत	कथ	कद	कध	कन
ख्र	खट	खठ	खड	खढ	खण	खत	खथ	खद	खध	खन
ग्र	गट	गठ	गड	गढ	गण	गत	गथ	गद	गध	गन
ज़्र	ज़ट	ज़ठ	ज़ड	ज़ढ	ज़ण	ज़त	ज़थ	ज़द	ज़ध	ज़न
फ्र	फट	फठ	फड	फढ	फण	फत	फथ	फद	फध	फन
ड्र	डट	डठ	डड	डढ	डण	डत	डथ	डद	डध	डन
ढ्र	ढट	ढठ	ढड	ढढ	ढण	ढत	ढथ	ढद	ढध	ढन

CHC SET 3:

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

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

ब्	ब्	ब्फ	ब्ब	ब्भ	ब्म	ब्य	ब्र	ब्ल	ब्ब	ब्श	ब्ष	ब्स	ब्ह
भ्	भ्प	भ्फ	भ्ब	भ्भ	भ्म	भ्य	भ्र	भ्ल	भ्व	भ्श	भ्ष	भ्स	भ्ह
म्	म्प	म्फ	म्ब	म्भ	म्म	म्य	म्र	म्ल	म्व	मश	म्ष	मस	मह
य्	य्प	य्फ	य्ब	य्भ	य्म	य्य	ग्र	ग्ल	ग्व	गश	ग्ष	गस	गह
र्	र्प	र्फ	र्ब	र्भ	र्म	र्य	र्र	र्ल	र्व	र्श	र्ष	र्स	र्ह
ल्	ल्प	ल्फ	ल्ब	ल्भ	ल्म	ल्य	ल्र	ल्ल	ल्व	लश	ल्ष	ल्स	ल्ह
व्	व्प	व्फ	व्ब	व्भ	व्म	व्य	व्र	व्ल	व्व	व्श	व्ष	व्स	व्ह
श्	श्प	श्फ	श्ब	श्भ	श्म	श्य	श्र	शल	श्व	शश	श्ष	शस	शह
ष्	ष्प	ष्फ	ष्ब	ष्भ	ष्म	ष्य	ष्र	ष्ल	ष्व	षश	ष्ष	ष्स	ष्ह
स्	स्प	स्फ	स्ब	स्भ	स्म	स्य	स्र	स्ल	स्व	सश	स्ष	स्स	सह
ह्	ह्प	ह्फ	ह्ब	ह्भ	ह्म	ह्य	ह्र	ह्ल	ह्व	ह्श	ह्ष	ह्स	ह्ह
क्	क्प	क्फ	क्ब	क्भ	क्म	क्य	क्र	क्ल	क्व	क्श	क्ष	क्स	क्ह
ख्	ख्प	ख्फ	ख्ब	ख्भ	ख्म	ख्य	ख्र	ख्ल	ख्व	खश	ख्ष	ख्स	ख्ह
ग्	ग्प	ग्फ	ग्ब	ग्भ	ग्म	ग्य	ग्र	ग्ल	ग्व	गश	ग्ष	गस	गह
ज्	ज्प	ज्फ	ज्ब	ज्भ	ज्म	ज्य	ज्र	ज्ल	ज्व	जश	ज्ष	जस	ज्ह
फ्	फ्प	फ्फ	फ्ब	फ्भ	फ्म	फ्य	फ्र	फ्ल	फ्व	फश	फ्ष	फ्स	फ्ह
ड्	ड्प	ड्फ	ड्ब	ड्भ	ड्म	ड्य	ड्र	ड्ल	ड्व	डश	ड्ष	ड्स	ड्ह
ढ्	ढ्प	ढ्फ	ढ्ब	ढ्भ	ढ्म	ढ्य	ढ्र	ढ्ल	ढ्व	ढश	ढ्ष	ढ्स	ढ्ह

CHC SET 4:

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

CHC Set 4							
	क्र	ख	ग	ज	फ	ड	ढ
क्	क्क्र	क्ख	क्ग	क्ज	क्फ	क्ड	क्ढ
ख्	खक्र	खख	खग	खज	खफ	खड	खढ
ग्	गक्र	गख	गग	गज	गफ	गड	गढ
घ्	घक्र	घख	घग	घज	घफ	घड	घढ
ङ्	ङक्र	ङख	ङग	ङज	ङफ	ङड	ङढ
च्	चक्र	चख	चग	चज	चफ	चड	चढ
छ्	छक्र	छख	छग	छज	छफ	छड	छढ
ज्	जक्र	जख	जग	जज	जफ	जड	जढ
झ्	झक्र	झख	झग	झज	झफ	झड	झढ
ञ्	ञक्र	ञख	ञग	ञज	ञफ	ञड	ञढ
ट्	टक्र	टख	टग	टज	टफ	टड	टढ
ठ्	ठक्र	ठख	ठग	ठज	ठफ	ठड	ठढ
ड्	डक्र	डख	डग	डज	डफ	डड	डढ
ढ्	ढक्र	ढख	ढग	ढज	ढफ	ढड	ढढ
ण्	णक्र	णख	णग	णज	णफ	णड	णढ
त्	त्क्र	त्ख	त्ग	त्ज	त्फ	त्ड	त्ढ
थ्	थक्र	थख	थग	थज	थफ	थड	थढ
द्	दक्र	दख	दग	दज	दफ	दड	दढ
ध्	धक्र	धख	धग	धज	धफ	धड	धढ
न्	न्क्र	न्ख	न्ग	न्ज	न्फ	न्ड	न्ढ
प्	पक्र	पख	पग	पज	पफ	पड	पढ
फ्	फक्र	फख	फग	फज	फफ	फड	फढ

ब्	ब्क	ब्ख	बा	ब्ज	बफ	ब्ड	ब्ढ
भ्	भ्क	भ्ख	भा	भ्ज	भफ	भड	भढ
म्	म्क	म्ख	मा	म्ज	मफ	मड	मढ
य्	य्क	य्ख	या	य्ज	यफ	यड	यढ
र्	र्क	र्ख	र्ग	र्ज	र्फ	र्ड	र्ढ
ल्	ल्क	ल्ख	ला	ल्ज	लफ	लड	लढ
व्	व्क	व्ख	वा	व्ज	वफ	वड	वढ
श्	श्क	श्ख	शा	श्ज	श्फ	श्ड	श्ढ
ष्	ष्क	ष्ख	षा	ष्ज	ष्फ	ष्ड	ष्ढ
स्	स्क	स्ख	सा	स्ज	स्फ	स्ड	स्ढ
ह्	ह्क	ह्ख	हा	ह्ज	हफ	हड	हढ
क्	क्क	क्ख	का	क्ज	क्फ	क्ड	क्ढ
ख्	ख्क	ख्ख	खा	ख्ज	खफ	खड	खढ
ग्	गक	गख	गा	गज	गफ	गड	गढ
ज्	ज्क	ज्ख	जा	ज्ज	जफ	जड	जढ
फ्	फ्क	फ्ख	फा	फ्ज	फफ	फड	फढ
ड्	ड्क	ड्ख	डा	ड्ज	डफ	डड	डढ
ढ्	ढ्क	ढ्ख	ढा	ढ्ज	ढफ	ढड	ढढ

7.3.3.2 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.³¹

CHCHC	Example
क्क्ष	दृक्क्षेप
क्त्र	इलेक्ट्रोन
क्त्स	स्पैक्ट्स
क्त्य	युक्त्याभास
क्त्र	पंचवक्त्र
क्त्व	पृथक्त्व
क्प्र	वाक्प्रतोद
क्त्य	शौक्ल्य
क्ष्ण	तीक्ष्ण
क्ष्म	भाग्यलक्ष्मी
क्ष्य	अनपेक्ष्य
क्ष्व	इक्ष्वाकु
क्स्ड	फिक्स्ड
क्स्त	वाक्स्तंभ
क्स्व	वाक्स्वातंत्र्य

³¹ Quite a few examples are culled from Bahri and Nalanda Dictionaries. Others are taken from the CIIL corpus.

ख्ख	ख्खवाजा
ग्ग	वणिग्गाम
ग्ज्य	प्राग्ज्योतिष
ग्द्व	प्राग्द्वार
ग्न्य	अग्न्याधान
ग्भ्र	दिग्भ्रम
ग्व्य	वाग्व्यवहार
क्ख्व	उक्ख्वास
ज्ज्य	उज्ज्वल
त्क्ल	उत्क्लेश
त्क्ष	तात्क्षणिक
त्थ	सुप्तोत्थित
त्त्व	तत्त्व
त्प्र	उत्प्रेरक
त्प्ल	उत्प्लव
त्स्य	स्थानमहात्स्य
त्र्य	वैचित्र्य
त्स्थ	तत्स्थानीय
त्स्न	जोत्स्ना
त्स्प	ईषत्स्पृष्ट
त्स्प	हत्स्पंदों
त्स्फ	उत्स्फूर्त
त्स्य	मत्स्य
त्स्व	तत्स्वरूप
द्ग्र	उद्ग्रीव

द्ज्ञ	सद्ज्ञान
द्द्व	एतद्द्वारा
द्दृष	नामोद्दिष्ट
द्भ्र	उद्भ्रम
द्रय	दारिद्र्य
द्व्य	सद्व्यवहारों
न्क्व	इन्क्वायरी
न्ट्र	इन्ट्रैन्ट्रियाँ
न्ट्स	स्टूडन्ट्स
न्त्य	अचिन्त्य
न्त्र	उप-मन्त्री
न्द्य	अनिन्द्य
न्द्र	अंतरिक्षकेन्द्र
न्द्र	उपेन्द्रनाथ
न्ध्य	फलवन्ध्य
न्त्य	वैभिन्न्य
न्क्र	कान्क्रेंस
प्य	प्राप्त्याशा
फ्ट्स	एयरक्राफ्ट्स
ब्ज्य	कौब्ज्य ³²
म्प्र	सम्प्राज
क्य	सक्यूलर
क्र	अंतर्क्रिया
कर्स	मार्क्सवाद

³² Sanskrit loan.: A hump-backed bull cf.

ख्य	मौख्य
ग्य	वर्ग्य
ग्र	अंतर्ग्रस्त
घ्य	अर्घ्य
र्छ	मूर्च्छित
र्च्य	अर्च्य
र्ज्ञ	अंतर्ज्ञान
ज्य	वर्ज्य
र्द्र	पोर्द्रेट
टर्स	स्पोर्ट्समैन
र्ण्य	वर्ण्य
र्त्य	अतिमर्त्य
र्त्र	सार्त्र
र्त्स	वर्त्स
र्थ्य	सामर्थ्यो
र्ध	आर्ध
र्द्र	आर्द्र
र्द्व	अंतर्द्वद्व 33
र्ध	पूर्वार्ध
र्ध्य	अंतर्ध्यान
र्ध्व	उर्ध्व
र्ष्य	समर्ष्य
र्प्र	अंतर्प्रादेशिक

³³ अंतर्द्वद्व . मनस्ताप - mental agony Arvind Lexicon.

फर्य	कफर्यू
भ्र	निभ्रत
मर्य	हमर्य
लर्ड	फ्रीवलर्ड
लर्स	चालर्स
वर्य	बहिर्यसनी
व्र	निव्रण
शर्व	पारशर्व
र्ण	कार्ण
र्ष	ईर्ष्या
स्ट	फ़स्ट
लक्य	याज्ञवल्क्य
ल्ट्र	पोल्ट्रीफार्म
ल्ट्र	चिल्ट्रन
लप्य	प्रकल्प्य
लभ्य	प्रागल्भ्य
श्व	पारिपाश्वक
ष्क्र	निष्क्रिय

ष्व	निष्वत्थ
ष्ट्य	उक्तिवैशिष्ट्य
ष्ट्र	अंतरराष्ट्रीय
ष्ट्य	ओष्ट्य
ष्य	वैतृष्य
ष्र	निष्रयोजन
स्क्र	स्क्रिप्ट
स्वव	मस्ववा
स्वव	स्ववाङ्मन
स्ट्र	इंडस्ट्रियल
स्त्य	अगस्त्य
स्त्र	अधोवस्त्र
स्त्व	अंतस्त्वचा
स्थ्य	स्वास्थ्य
स्प्र	स्प्रिंग
स्प्ल	एस्प्लेनेड
द्व	सद्व्रत

7.3.3.3.CHCHCHC (Combination of four Consonants)

These combinations are rare and found only in the following case³⁴:

स्त्र्य स्त्रयाजीव

7.3.3.4 CHCHCHCHC (Combination of five Consonants)

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

³⁴ The examples are culled from Bahri and Nalanda Dictionaries.

7.3.4 Valid Combinations And Invalid Combinations

As mentioned in the *Introduction* as well as Section 1. *End Users of the “Devanāgarī 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.

7.3.5 The Collation Order of Hindi.

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

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

In the case of Hindi the following is the traditional sort order as determined by the experts and used in dictionaries and may have unique variations across dictionaries.

अ	आ	इ	ई	उ	ऊ	ऋ	ए	ऐ ³⁵	ओ	औ
क	ख	ग	घ	ङ	च	छ	ज	झ	ञ	ट
ठ	ड	ढ	ण	त	थ	द	ध	न	प	फ
ब	भ	म	य	र	ल	व	श	ष	स	ह

The order as given below is pertinent to sorting by a computer program and is compliant with CLDR as laid down by Unicode³⁶ and is recommended for use.

ॠ ॡ ॢ ॣ । ॥ ० १ २ ३ ४ ५ ६ ७ ८ ९ ॰ ॱ ॲ ॳ ॴ ॵ ॶ ॷ ॸ ॹ ॺ ॻ ॼ ॽ ॾ ॿ ॠ ॡ ॢ ॣ । ॥ ० १ २ ३ ४ ५ ६ ७ ८ ९ ॰ ॱ ॲ ॳ ॴ ॵ ॶ ॷ ॸ ॹ ॺ ॻ ॼ ॽ ॾ ॿ

³⁵ ऐ and औ are normally not listed in traditional dictionaries, but are included as digital dictionaries do include these characters.

³⁶ As per CLDR ver. 25 <http://cldr.unicode.org/index/downloads/cldr-25>

Following is the example of sort order for consonant “क³⁷,”

क	कँ	कं	कः	का	कि	की	कु	कू	कृ	के	कै
कै	को	कौ	कौ	क्	क्र						

³⁷ क is placed at the top of the sort table since it constitutes an Indic Orthographic Syllable by itself. The other combinations listed after the consonant constitute matra or diacritic combinations.

8. REFERENCES

1. <http://www.unicode.org>
2. <http://cldr.unicode.org/>
3. INDIAN SCRIPT CODE FOR INFORMATION INTERCHANGE - ISCII IS 13194. 1991. NEW DELHI
4. केंद्रीय हिंदी निदेशालय: देवनागरी लिपि तथा हिंदी वर्तनी का मानकीकरण नई दिल्ली. 2010
5. Omkar N Koul *Modern Hindi Grammar*. Hyattsville: Dunwoody Press, 2008, pp. x+318
6. Yamuna Kachru. *Hindi*. London Oriental and African Language Library, 12 2006. xxii, pp. 309
7. Kalyan Kale, Anjali Soman. *Learning Marathi* 1st ed. Shri Vishakha Prakashan, 1986.

9. ANNEXURES

Annexure 1 : Names of experts who contributed in evolving the draft standard.

Shapes in this documents are based on CHD book.

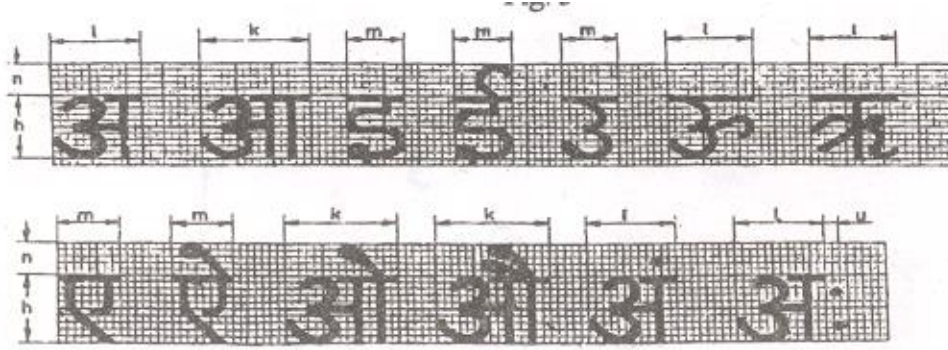
Contributors :

S.No.	Name	Organization	Email
1.	Ms. Swaran Lata	DeitY	slata@mit.gov.in
2.	Sh. Manoj Jain	DeitY	mjain@mit.gov.in
3.	Sh. Vijay Kumar	DeitY	vkumar@mit.gov.in
4.	Sh. Bharat Gupta	DeitY	bharatg@deity.gov.in
5.	Hindi language expert	CIIL	
6.	Dr. Raiomond Doctor	C-DAC, GIST, Pune	rdoctor@cdac.in
7.	Sh. M D Kulkarni	C-DAC, GIST , Pune	mdk@cdac.in
8.	Mr. Prabhakar Pandey	C-DAC, GIST , Pune	prabhakarp@cdac.in
9.	Dr. Sudhir Mishra	C-DAC, GIST , Pune	smishra@cdac.in
10.	Mr. Tukaram Patil	Department of Hindi, University of Pune	
11.	Dr. Mukul Kumar Sinha	Expert Software consultants	mukuls@yahoo.co.in
12.	Prof. Panchanan Mohanty	University of Hyderabad	panchananmotianty@gmail.com
13.	Prof. B. Mallikarjun	Central University or Karnataka	mallikarjun56@gmail.com
14.	Prof. Ghanashyam Nepal	University of North Bengal	nepalgs@gmail.com
15.	Prof. Hidam Dolen Singh	NERLC	dikebgu@gmail.com
16.	Dr. Bhaskarjyoti Sarma	Dibrugarh University	bhasdu08@gmail.com
17.	Sh . J.Kumar	NIC Delhi	kendolji@nic.in
18.	Prof. Bisweswar Basumatary	Bodo Sahitya Sabha	bssjb@rediffmail.com
19.	Sh. Deepak Pandey	CHD	dkp410@gmail.com
20.	Ms. Nutan Pandey	CHD	ilove.india2008@indiatimes.com
21.	Ms. Richa	LDC-IL, CIIL Mysore	rsrishti@gmail.com
22.	Prof. Girish Nath Jha	JNU	girishjha@gmail.com
23.	Sh Pravin Satpute	Redhat	psatpute@redhat.com
24.	Prof K. K.Goswami		
25.	Dr. Saurobh Gupta	NIC	Saurabh.gupta@nic.in
26.	Sh R.K. Pathak	CTB, DOL, MHA	Rakesh.pathak@nic.in
27.	Sh. Kewal Krishan	NIC, D/O official language	kewal.krishan@nic.in
28.	Sh. Harinder Kumar	DOL, MHA	dol-tech@nic.in

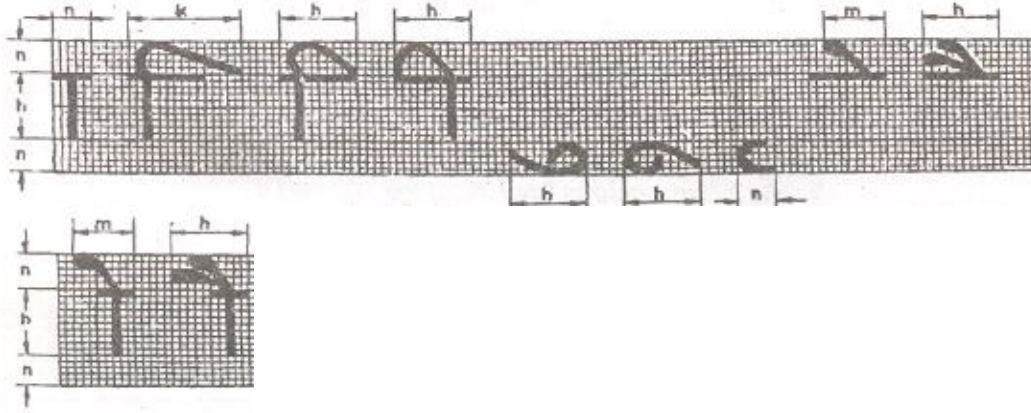
29.	Prof. N Pramodni Devi	Manipur University	npdini@yahoo.co.in
30.	Prof. Shikhar Kumar Sharma	Gauhati University	sks001@gmail.com
31.	Prof. V R Jagannathan	Retd	vrj.nathan@gmail.com
32.	Prof. Pushpak Bhattatacharya	IITB, Mumbai	pb@cse.iitb.ac.in
33.	Sh. Abhijit Bhattacharjee	Luna Ergonomics	abhijit@paninikeyped.com

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

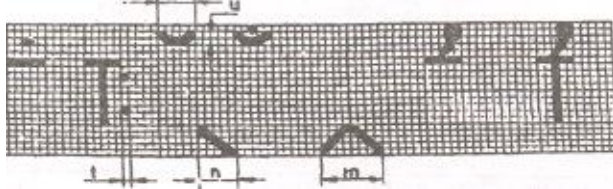
Hindi Vowels



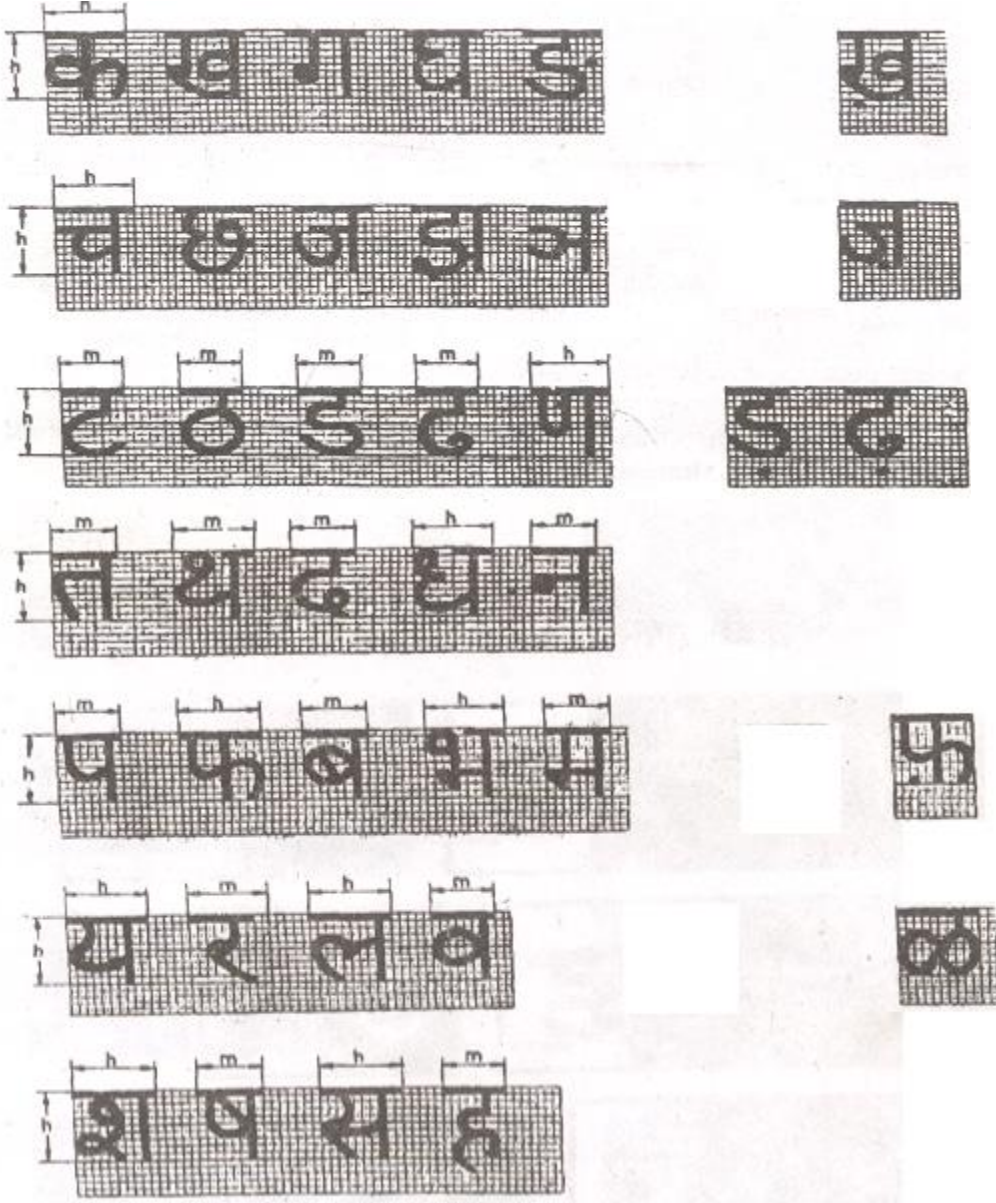
Hindi Vowel Modifiers: Mātrās



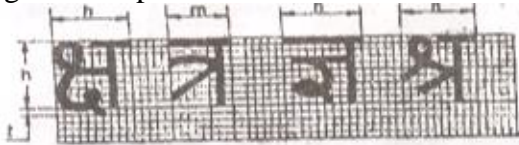
Hindi Diacritics and Rafar(Repha) Rakar



Hindi Consonants



Ligatural Representations



Annexure 3: Unicode Table of Devanāgarī as per UNICODE 7.0³⁸

	090	091	092	093	094	095	096	097
0	ॐ	ऐ	ठ	र	ी	ॐ	ॠ	ॡ
1	ॐ	ऑ	ड	र	ॐ	ॐ	ॠ	ॡ
2	ॐ	ओ	ढ	ल	ॐ	ॐ	ॠ	ॡ
3	ॐ	ओ	ण	ळ	ॐ	ॐ	ॠ	ॡ
4	अ	औ	त	ळ	ॐ	ॐ	ॠ	ॡ
5	अ	क	थ	व	ॐ	ॐ	ॠ	ॡ
6	आ	ख	द	श	ॐ	ॐ	ॠ	ॡ
7	इ	ग	ध	ष	ॐ	ॐ	ॠ	ॡ
8	ई	घ	न	स	ॐ	ॐ	ॠ	ॡ
9	उ	ङ	न	ह	ॐ	ॐ	ॠ	ॡ
A	ऊ	च	प	ं	ो	ग	४	य
B	ॠ	छ	फ	ी	ो	ज	५	ग
C	ॠ	ज	ब	॑	ौ	ड़	६	ज
D	ँ	झ	भ	ऽ	॑	ढ़	७	१
E	ऐ	ज	म	ा	ि	फ़	८	ड
F	ए	ट	य	ि	ौ	य़	९	ब

The Unicode Standard 7.0, Copyright © 1991-2014 Unicode, Inc. All rights reserved.

³⁸ The Unicode chart provided is for version 7.0. The latest Devanagari chart can be seen at <http://unicode.org/charts/>

Annexure 4: Note On Ambiguous Use Of Anuswara (Anunasika) With Superscript Vowel Signs In Hindi³⁹

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/ सां/ पुं can stand for kəm, kən, kəṇ, kəṇ and kəṇ/ sam, san, saṇ, saṇ and saṇ/ pum, pun, puṇ, puṇ and puṇ, respectively.

The ‘Candrabindu’ stands for anuswāra on subscripted vowel mātrās: thus कँ/ साँ/ पुँ stand for only kṣ̃/ sã/ /pũ/ 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 कौ stands for both anuswāra (kō as in कौपल ‘bud’) and anunāsika (kom, kon, koṇ, koṇ, koṇ as in कौडा, कौकण etc.). The examples can be given for all the four vowel mātrās represented with superscript diacritics, e.g. सिंध ‘Sindh’ सिंचाई ‘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 Candrabindu 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.

³⁹ Dr Pramod Pandey, Centre for Linguistics, J.N.U., New Delhi. (slightly modified version)

PART B

RECOMMENDATIONS FOR DEVELOPERS

Language specific issues are complex and the developers require a ready reckoner for reference. Although this section may not be considered an integral part of Script Grammar per se, yet the following guidelines are laid down for the ease of developers and are documented as a separate section in the Script Grammar. Indian scripts are derived from Brahmi and are abugidas (alphasyllabaries); Abugidas have letters to represent consonant sounds; vowel sounds and vowel sounds are also shown with diacritics (modifiers to the consonant letters). Each consonant letter carries an inherent vowel i.e. अ. The chapter 6 of Unicode standard describes the various writing systems in detail.

Syllable is an element of a syllabary and is the basic unit of articulation that corresponds to a pulmonary pulse. Hence the knowledge of the Indic Orthographic Syllable is a must. Thus, there is need to evolve a basic formalism of Indic Orthographic Syllable, which facilitates implementation of correct cursor movement etc. This definition will also ensure correct implementation of syllable in various applications. 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.

To help the developers address, these issues, part B of Script Grammar has been divided into two major sections: viz. Indic Orthographic Syllable and Linguistic Issues

In the first section, all issues afferent to the Orthographic Syllable are handled, which is technical in nature and addresses the Indic syllable and provides a general overview of the same. This is followed by an overview of the formalism underlying the Orthographic Syllable and finally issues derived from the Orthographic Syllable such as cursor movement, normalisation etc. This part of the document is more a compendium and reference to the relevant resources has been provided as footnotes or in the main body of the text.

In the second section, 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. 2010.

Section I

1. TECHNICAL ISSUES

This part starts off with 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 permits more than one way of inputting a given character, Normalization is of prime importance.

1.1. DEVANĀGARĪ SCRIPT AN OVERVIEW⁴⁰

This over-view is a linguistic introduction to Devanāgarī. It starts off with the historical evolution of Devanāgarī and in Section 1.3 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

Devanāgarī 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 Devanāgarī 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 Devanāgarī involved intermediate forms, common to other scripts such as Gupta and Śāradā in the north and Grantha and Kadamba in the South. Devanāgarī 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 Devanāgarī 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.

⁴⁰⁴⁰Extracted 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/Devanāgarī+VIP+Team+Issues+Report+1.0.docx?version=1&modificationDate=1315604624000>.

1.3. ENCODING PRINCIPLES

Chapter 9 of Unicode 6.1⁴¹ clearly lays down the encoding principles underlying Devanāgarī script. The relevant pages 211-215 give an overview of the basic structure of Devanāgarī 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 the document.

1.4. THE FUNDAMENTAL UNIT: SYLLABLE

The syllable is a basic unit of speech studied on both the phonetic and phonological levels of analysis. It is very easy for a native language speaker to count the number of syllables in a sequence however Orthographic syllable required for text processing need not correspond exactly with a phonological syllable.

Indic Orthographic Syllable has been defined as a computational formalism using Augmented Backus-Naur Formalism (ABNF) - refer 6.2 of Part A.

1.5. UNICODE TEXT SEGMENTATION UAX #29

The notion of Indic Orthographic Syllable is essential to this topic which defines “*guidelines for determining default boundaries between certain significant text elements: user-perceived characters, words, and sentences*”⁴².

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

This process whereby boundaries are delimited results in segmentation. 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 for Indic Orthographic Syllable segmentation will illustrate the importance of segmentation.

Given a string such as:

⁴¹ <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 download the document for details. This applies to all Unicode documents which are protected by copyright laws.

⁴² Op cit.

मनमोहन ओष्ठ्य

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

म- न-मो- ह -न

C - C - Cv -C - C

ओ- ष्ट्य

V - C+H+C+H+C

नि - ष्क्रि - य

Cv – C+H+C+H+C+v - C

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

1.6. RENDERING

Closely allied to the issues studied above is the problem of rendering. Chapter on “South and Central Asia-I” of Unicode studies in depth the issue of rendering in Devanāgarī.

“Devanāgarī 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 Devanāgarī characters to differ from their nominal glyphs (used in the code charts). Additionally, a few Devanāgarī 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.”⁴³

The Chapter⁴⁴ 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 to prevent conjunct forms, generating explicit half consonants

⁴³<http://www.unicode.org/versions/Unicode6.1.0/ch09.pdf>, p. 212

⁴⁴ pp. 211-22

and most importantly

- Rendering Rules

Although the use of ZWJ/ZWNJ is pre-implied in the chapter, it will be treated separately in 1.7

1.7. ZWJ/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(द) U+0930(र) U+094D(्) U+200D(ZWJ) U+092F(य) U+093E(ा) (दर्या)

as well as

U+0926(द) U+0931(र) U+094D(्) U+092F(य) 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:

शक्ति U+0936(श) U+0915(क) U+094D(्) U+200D(ZWJ) U+0924(त) U+093F(ि)

अक्षय U+0905(अ) U+0915(क) U+094D(्) U+200D(ZWJ) U+0937(ष) U+092F(य)

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

RULE 2: 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(र) U+093E(ा) U+091C(ज) U+094D(्) U+200C(ZWNJ) U+0915(क)
U+092E(म) U+0932(ल)

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 ZWJ/ZWNJ पक्का

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.

1.8. CURSOR MOVEMENT& DELETION

1.8.1 Cursor movement

Cursor movement is an important issue in content creation. Given the complex nature of Brahmi based scripts as opposed to Perso-Arabic; 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 Devanāgarī in this instance, the cursor movement is slightly different as the following quote⁴⁵ 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

⁴⁵<http://msdn.microsoft.com/en-us/goglobal/bb688137.aspx>

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 Devanāgarī, the cursor will move across Indic Orthographic Syllables and not across the individual characters.

The Table below shows the basic difference between Devanāgarī and Latin

MOVEMENT	LATIN	DEVANĀGARĪ
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
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(र) U+0915(क) U+094D(्) U+200D(ZWJ) U+0924(त)

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

रक्त U+0930(र) U+0915(क) U+094D(्) U+200C(ZWNJ) U+0924(त)

1.8.2 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 Devanāgarī which has the Indic Orthographic Syllables 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 → 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.

रक्त → रक् → रक → 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

रक्त → क्त → त

1.9. NORMALIZATION

A major issue in Hindi is that of Normalization. Normalization is a major issue and is of such importance that the document Unicode Standard Annex #15⁴⁶ has undergone multiple changes.

In simple language Normalization takes place when a given 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 scripts⁴⁷. 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

क 0958 क 0915 093C

इ 095C इ 0921 093C

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

Normalized combination	Normalized form
क + ं 0915 + 093C	क 0958
ख + ं 0916 + 093C	ख 0959
ग + ं 0917 + 093C	ग 095A

⁴⁶<http://unicode.org/reports/tr15/>

⁴⁷<http://www.unicode.org/charts/normalization/>

ज + ँ 091C+ 093C	ज़ 095B
ड + ँ 0921 + 093C	ड़ 095C
ढ + ँ 0922 + 093C	ढ़ 095D

Table 5

In the case of browsers within the URL bar, a URL containing a word with Consonant+nukta is automatically normalised to its corresponding Akhand form. Thus:

बढ़.com U+092C(ब) U+095D(ढ़).com

is automatically normalized to

बढ़.com U+092C(ब) U+0922(ढ) U+093C(ं).com

Developers would do well to study the document to understand the complexity of the problem along with the FAQ on Normalization⁴⁸ as well as the Normalization, Sorting and Searching sections of chapter 5 of the Unicode Standard⁴⁹. Detection and Testing of Normalization forms is also provided in useful links⁵⁰. ICU also supports related, additional operations. Some of them are described in Unicode Technical Note #5 (Canonical Equivalence in Applications)⁵¹.

⁴⁸<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.

⁴⁹<http://www.unicode.org/versions/Unicode6.2.0/>

⁵⁰Normalization forms detection

[http://unicode.org/reports/tr15/#Detecting Normalization Forms](http://unicode.org/reports/tr15/#Detecting_Normalization_Forms),

Normalization test

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

⁵¹<http://www.unicode.org/notes/tn5/>

SECTION II

1. LINGUISTIC ISSUES

2.1. PRELIMINARIES

This section is based on the 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:

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 प्रातिपदिक से पृथक् लिखे जाएँ। जैसे :- राम ने, राम को, राम से, स्त्री का, स्त्री से, सेवा में आदि। सर्वनाम शब्दों में ये चिह्न प्रातिपदिक 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 (हाइफ़न) का विधान स्पष्टता के लिए किया गया है।

द्वंद्व समास में पदों के बीच हाइफ़न 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 हाइफ़न रखा जाए जैसे :- तुम-सा, चाकू-से तीखे।

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

इसी तरह यदि 'अ-नख' (बिना नख का) समस्त पद में हाइफ़न 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 हाइफ़न का प्रयोग किया जा सकता है। जैसे :- द्वि-अक्षर न कि द्व्यक्षर; द्वि-अर्थक न कि द्व्यर्थक, त्रि-अक्षर न कि त्र्यक्षर आदि।

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

Both Anuswar and Candrabindu shall be used in Hindi. However the usage of these two shall be governed by precise guide-lines which are stipulated at length and which 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 (शिरोबिंदुं)

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 सवर्गीय 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 :- वाङ्मय, अन्य, चिन्मय, उन्मुख आदि (वांमय, अंय, चिमय, उंमुख आदि रूप ग्राह्य नहीं होंगे)।

पंचम वर्ण यदि द्वित्व रूप में (साथ-साथ) आए तो पंचम वर्ण अनुस्वार में परिवर्तित नहीं होगा। जैसे :- अन्न, सम्मेलन, सम्मति आदि (अंन, संमेलन, संमति रूप ग्राह्य नहीं होंगे)। अंग्रेज़ी, उर्दू से गृहीत शब्दों में आधे वर्ण या अनुस्वार के भ्रम को दूर करने के लिए नासिक्य व्यंजन को पूरा लिखना अच्छा रहेगा। जैसे :- लिमका, तनखाह, तिनका, तमगा, कमसिन 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 Candrabindu.

The following are the guidelines governing the use of Candrabindu:

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, में, 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 नि दु the visarga is replaced by a doubling of the consonant as in दुःसाहास निःशब्द. These 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)।

दुःसाहास/दुस्साहास, निःशब्द/निश्शब्द के उभय रूप मान्य होंगे। इनमें द्वित्व वाले रूप को प्राथमिकता 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, प्रातःकाल 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 Chihna 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, ढ, द, ह the guidelines also amplify and provide recommendations for the use of the Halant.

(्) को हल् चिह्न कहा जाए न कि हलन्त। 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’ हलन्त शब्द कहा जाएगा क्योंकि यह शब्द व्यंजनांत है, स्वरांत नहीं।

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

तत्सम शब्दों का प्रयोग वांछनीय हो तो हलन्त 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 :- महान, विद्वान आदि; क्योंकि हिंदी में अब ‘महान’ से ‘महानता’ और विद्वान 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 cf. Section 3.10 of the document.

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

हिंदी में ऐ (ै), 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 ‘गवैया’, ‘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, ै; औ, ौ) का प्रयोग किया जाए। ‘गवय्या’, ‘कव्वा’ आदि संशोधनों की आवश्यकता नहीं है। अन्य उदाहरण हैं – भैया, सैयद, तैयार, हौवा आदि।

दक्षिण के अय्यर, नय्यर, रामय्या आदि व्यक्तिनामों को हिंदी उच्चारण के अनुसार ऐयर, नैयर, रामैया आदि न लिखा जाए, क्योंकि मूलभाषा में इसका उच्चारण भिन्न है।

अव्वल, कव्वाल, कव्वाली जैसे शब्द प्रचलित हैं। इन्हें लेखन में यथावत् रखा जाए।

संस्कृत के तत्सम शब्द ‘शय्या’, 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 मकानvillage house

Each case is handled below

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

h. Recommendations for usage of य व

Closely associated with the usage of ऐ & 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 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 यी, the former being the preferred choice. This applies as the guidelines explicitly state to cases given below:

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.

The guidelines of the CHD are reproduced below:

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. 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:

3.1. 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 (कलम, किला, दाग 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 : ख़ाना, राज : राज़, 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 —

नुक़सान, इनसान, इनकार, परवाह, बरबाद, मज़दूर, सरकार, फरमान, तनखाह, करवट, कुदरत, शिरकत, मरकज़, अक़बर, मुजरिम, परवाना, मुग़दर, मसख़रा, इलज़ाम, फ़िलहाल, फ़ुरसत, मनज़र, (दृश्य), इनक़लाब, आसमान, तकलीफ़, मरहम, दरकार, किरदार, शबनम, इज़लास, इक़बाल, इनतहा, तनहा, तनहाई, असल, असली, असलियत, ग़लत, ग़लती, नक़ल, नकली, नक़लची, वरना, शरबत, गरदन, दफ़्तर, सरक़स, बरतन, बिलकुल, अक़सर, रफ़्तार, गरम, गरमी, शरम, शरमाना, सरदी आदि। (इनसान, इनकार, तनखाह, मनज़र, इनक़लाब, इनतहा, तनहा, तनहाई आदि को किसी भी स्थिति में इंसान, इंकार, तंखाह, मंज़र, इंक़लाब, 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:

फ़ुर्ती, फ़ुर्तीला, शिक़स्त, दुरुस्त, बुज़ुर्ग, गिरफ़्त, गिरफ़्तार, रुख़्सत, ज़िंदा, कर्ज़, खर्च, ज़िंदगी, इंतक़ाम, इंतज़ार, इल्तिज़ा, अशफ़ी, इम्तहान, तिलिस्म, पुख़्ता, दरख़्त, तरार, तश्तरी, लश्कर, जिन्स, मुंशी, नक्शा, कब्ज़, लफ़्ज़, शुक्र, शुक्रिया अख़्तियार, इत्तेफ़ाक़, इश्तहार, बख़्शीश, अक्ल, शक्ल, सख़्त, सख़्ती, ग़श्त, किस्त, किस्म, मुश्क, इश्क, जश्न, कश्मीर, पश्मीना, जुल्म, जुर्म, जिस्म, रस्म, रस्मी, हश्म, फ़िक्क, मुल्क, इल्म, फ़र्ज, फ़र्जी, इर्द-गिर्द, गोश्त, रास्ता, नाश्ता, शख़्स, दर्ज़, दर्जी, सुख़्र, सुख़्री, गुर्दा, बर्फ़, जल्द, जल्दी, चश्मा, भिश्ती, सर्द, उम्र, कुश्ती, चुस्त, चुस्की, फ़र्क, 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, फ़ुरसत/फ़ुर्सत, 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, दुमंज़िला, 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 उर्दू 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 निस्वत (के लिए) आदि।

मुख़ालफ़त (विरोध), ज़ाहिर (स्पष्ट), हमला (आक्रमण), कामयाबी (सफलता), मुहैया कराना (उपलब्ध कराना), पेश करना (प्रस्तुत करना), मकसद (उद्देश्य), हद (सीमा), जिहादी (धर्म युद्ध करने वाला), शुरुआत (आरंभ), कोताही (लापरवाही), गौरतलब (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)

कागज़ = बहुवचन कागज़ों (में) या कागज़ात में (कागज़ातों में नहीं)

किताब = बहुवचन किताबें (किताबात नहीं)

किस्म = किस्में (अक्साम नहीं)

3.2. 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।
 बांग्ला शब्द - रसगुल्ला, चमचम, उपन्यास, गल्प, आपत्ति आदि।
 पंजाबी शब्द - सिकख, 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 - हड़ताल, 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।

2.3.3 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, फ्रीस, पिन, पेट्रोल, पुलिस, पेंसिल, 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, कागज़, कानून, अल्लाह आदि।

फ़ारसी - कमर, कम, ख़ाक, गुम, वापस, ख़ुदा 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।

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

स्पेनी - सिगार, सिगरेट, कार्क, पिउन आदि।

जर्मन - ट्रेन, सेमिनार आदि।

जापानी - रिक्शा, जूड़ो आदि।

रूसी - सोवियत, स्पुतनिक, जार आदि।

इतालियन - लॉटरी, कार्टून, 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 इन्वर्टे 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:

अंग्रेज़ी शब्दों का लिप्यंतरण करते समय शब्दांत में दीर्घ ‘ई’ तथा ‘ऊ’ का ही प्रयोग होगा। जैसे :-

dye	डाई	waterloo	वाटरलू
pie	पाई	shoe	शू
City	सिटी	pharmacy	फार्मेसी
daddy	डैडी	company	कंपनी
luxury	लग्ज़री	blue	ब्लू
Key	की	waterloo	

Insofar as words of Latin and Slav 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: तोल्सतोय/तोल्सतोई (Tolstoy)
दूप्ले (Duplex) स्पेनी: दान किहोते (डॉनक्विज़ोट)
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. PUNCTUATIONS⁵²

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 software does 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 DEVANĀGARĪ SCRIPT BEHAVIOUR FOR HINDI:

Devanāgarī Name	Name of the marker	Marker Shape
"स्टॉप" ⁵³	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

⁵²Section 4.0 of the document. The correct Unicode values of punctuations are provided

⁵³<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 DEVANĀGARĪ ABBREVIATION SIGN	◌
पूर्ण विराम	U+0964 DEVANĀGARĪ DANDA Devanāgarī Purna Viram	
दीर्घ विराम	U+0965 DEVANĀGARĪ DOUBLE DANDA Devanāgarī Deergha Viram	
हंसपद त्रुटिबोधक	/ Sign of left/Add word	↙

4.1 Full Stop: (।) or पूर्ण विराम⁵⁴

- 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 उपशीर्षकों के अंत में पूर्ण विराम न दिया जाए। जैसे :- प्राणी तीन तरह के हैं : (क) थलचर (ख) जलचर (ग) नभचर

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-

⁵⁴The term as used by the Central Hindi Directorate refers to the purnaviram and not the English full stop.

- 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, घृणा, 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 स्वीकृती/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 शत्रु 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। अंग्रेजी 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 श्रृंखला 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 मुनष्य 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 गाँधी 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 कोई 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, ख़ैर** 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 बलिए।

वास्तव में, उसकी कार्य-पद्धति प्रशंसनीय है।

ख़ैर, 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 जरूर 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 डाल 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 के 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 अंग्रेज़ी के अपॉस्ट्रॉफी के लिए किया जाता है। अंग्रेज़ी 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 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 आज़ा 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 (शांतिपूर्वक) - “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^० 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 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 टिप्पणीयाँ/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 तात्काल याद न आ रहा हो तो लोप चिह्न प्रयुक्त होता है और जब बात छूटी हुई 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 (.) (°)

संक्षिप्ति बिंदु (°) का प्रयोग शून्य (.) की भाँति भी होता है। जैसे :- एम. ए. ~ एम. ए. (मास्टर ऑफ़ आर्ट्स), रा.च. प्रसाद ~ रा.च. प्रसाद (राम चंद्र प्रसाद), पं. ~ पं. (पंडित), 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 व्यवस्था के / 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 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 the spellings of Cardinal Numbers are 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	सैंतालीस	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	

हज़ार, लाख, करोड़, अरब (सौ करोड़ नहीं), साठ अरब (छह हज़ार करोड़ नहीं)
पाँच खरब (पचास हज़ार 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 दोन 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
	स्वर	1.1
	स्वर की परिभाषा	1.1.1
	स्वर भेद	1.1.2
	रचना के अनुसार	1.1.2.1
	मूल (एकस्वरक)	1.1.2.1.1
	दीर्घीकृत	1.1.2.1.2
	दीर्घ	1.1.2.1.2. 1
	प्लुत	1.1.2.1.2. 2
	ध्यक्षर	1.1.2.1.3
	अनुनासिकता के आधार पर	1.1.2.2
	मौखिक/निरनुनासिक	1.1.2.2.1
	अनुनासिक	1.1.2.2.2
	व्यंजन	1.2
II.	शब्दविचार	2.0
III.	वाक्यविचार	3.0
IV.	रचना	4.0

REFERENCES

Indic Layout Requirements- This document describes the basic requirements for Indian Languages layout for display purpose. It discusses some of the major layout requirements in first letter pseudo-element, vertical arrangements of characters, letter spacing, text segmentation, line breaking and collation rules in Indic languages. Please refer following link for more information:

<http://www.w3.org/International/docs/indic-layout/>