L2/02/326

9.1 Devanagari

Devanagari: U+090[illegible]

The Devanagari script is used fo[illegible]odern historical derivative, Hindi. Extensions to Dev[illegible]ed languages of India (such as Marathi) and of Nepal [illegible] script is used to write ~~the following languages: Awad[illegible]raj Bhasha, Chhattisgarhi, Garhwali, Gondi (Betul[illegible] Harauti, Ho, Jaipuri, Kachchhi, Kanauji, Konkani, Ku[illegible]ri, Mundari, Newari, Palpa, and Santali~~.

All other Indic scripts as well a[illegible], the Tibetan script, and the Southeast Asian scripts (Thai, Lao, [illegible], are historically connected with the Devanagari script as descendants of the ancient Brahmi script. The entire family of scripts shares a large number of structural features.

The principles of the Indic scripts are covered in some detail in this introduction to the Devanagari script. The remaining introductions to the Indic scripts are abbreviated but highlight any differences from Devanagari where appropriate.

Standards. The Devanagari block of the Unicode Standard is based on ISCII-1988 (Indian Standard Code for Information Interchange). The ISCII standard of 1988 differs from and is an update of earlier ISCII standards issued in 1983 and 1986.

The Unicode Standard encodes Devanagari characters in the same relative position as those coded in positions A0–F4$_{16}$ in the ISCII-1988 standard. The same character code layout is followed for eight other Indic scripts in the Unicode Standard: Bengali, Gurmukhi, Gujarati, Oriya, Tamil, Telugu, Kannada, and Malayalam. This parallel code layout emphasizes the structural similarities of the Brahmi scripts and follows the stated intention of the Indian coding standards to enable one-to-one mappings between analogous coding positions in different scripts in the family. Sinhala, Thai, Lao, Khmer, and Myanmar depart to a greater extent from the Devanagari structural pattern, so the Unicode Standard does not attempt to provide any direct mappings for these scripts to the Devanagari order.

In November 1991, at the time *The Unicode Standard, Version 1.0*, was published, the Bureau of Indian Standards published a new version of ISCII in Indian Standard (IS) 13194:1991. This new version partially modified the layout and repertoire of the ISCII-1988 standard. Because of these events, the Unicode Standard does not precisely follow the layout of the current version of ISCII. Nevertheless, the Unicode Standard remains a superset of the ISCII-1991 repertoire except for a number of new Vedic extension characters defined in IS 13194:1991 *Annex G—Extended Character Set for Vedic*. Modern, non-Vedic texts encoded with ISCII-1991 may be automatically converted to Unicode code points and back to their original encoding without loss of information.

Encoding Principles. The writing systems that employ Devanagari and other Indic scripts constitute a cross between syllabic writing systems and phonemic writing systems (alphabets). The effective unit of these writing systems is the orthographic syllable, consisting of a consonant and vowel (CV) core and, optionally, one or more preceding consonants, with a canonical structure of $((C)C)CV$. The orthographic syllable need not correspond exactly with a phonological syllable, especially when a consonant cluster is involved, but the writing system is built on phonological principles and tends to correspond quite closely to pronunciation.

The orthographic syllable is built up of alphabetic pieces, the actual letters of the Devanagari script. These pieces consist of three distinct character types: consonant letters, independent vowels, and dependent vowel signs. In a text sequence, these characters are stored in logical (phonetic) order.

Principles of the Script

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

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

Consonant Letters. Each consonant letter represents a single consonantal sound but also has the peculiarity of having an *inherent vowel*, generally the short vowel /a/ in Devanagari and the other Indic scripts. Thus U+0915 DEVANAGARI LETTER KA represents not just /k/ but also /ka/. In the presence of a dependent vowel, however, the inherent vowel associated with a consonant letter is overridden by the dependent vowel.

Consonant letters may also be rendered as *half-forms*, which are presentation forms used to depict the initial consonant in consonant clusters. These half-forms do not have an inherent vowel. Their rendered forms in Devanagari often resemble the full consonant but are missing the vertical stem, which marks a syllabic core. (The stem glyph is graphically and historically related to the sign denoting the inherent /a/ vowel.)

Some Devanagari consonant letters have alternative presentation forms whose choice depends upon neighboring consonants. This variability is especially notable for U+0930 DEVANAGARI LETTER RA, which has numerous different forms, both as the initial element and as the final element of a consonant cluster. Only the nominal forms, rather than the contextual alternatives, are depicted in the code chart.

The traditional Sanskrit/Devanagari alphabetic encoding order for consonants follows articulatory phonetic principles, starting with velar consonants and moving forward to bilabial consonants, followed by liquids and then fricatives. ISCII and the Unicode Standard both observe this traditional order.

Independent Vowel Letters. The independent vowels in Devanagari are letters that stand on their own. The writing system treats independent vowels as orthographic *CV* syllables in which the consonant is null. The independent vowel letters are used to write syllables that start with a vowel.

Dependent Vowel Signs (Matras). The dependent vowels serve as the common manner of writing noninherent 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. A single consonant, or a consonant cluster, may have a dependent vowel applied to it to indicate the vowel quality of the syllable, when it is different from the inherent vowel. Explicit appearance of a dependent vowel in a syllable overrides the inherent vowel of a single consonant letter.

The greatest variation among different Indic scripts is found in the way that the dependent vowels are applied to base letterforms. Devanagari has a collection of nonspacing dependent vowel signs that may appear above or below a consonant letter, as well as spacing dependent vowel signs that may occur to the right or to the left of a consonant letter or consonant cluster. Other Indic scripts generally have one or more of these forms, but what is a nonspacing mark in one script may be a spacing mark in another. Also, some of the Indic scripts have single dependent vowels that are indicated by two or more glyph components—and those glyph components may *surround* a consonant letter both to the left and right or may occur both above and below it.

The Devanagari script has only one character denoting a left-side dependent vowel sign: U+093F DEVANAGARI VOWEL SIGN I. Other Indic scripts either have no such vowel signs (Telugu and Kannada) or include as many as three of these signs (Bengali, Tamil, and Malayalam).

A one-to-one correspondence exists between the independent vowels and the dependent vowel signs. Independent vowels are sometimes represented by a sequence consisting of the independent form of the vowel /a/ followed by a dependent vowel sign. For example, *Figure 9-1* illustrates this relationship (see the notation formally described in the "Rules for Rendering" later in this section).

Figure 9-1. Dependent Versus Independent Vowels

/a/ + Dependent Vowel		Independent Vowel
$A_n + I_{vs} \rightarrow I_{vs} + A_n$	=	I_n
अ + ि → अि	=	इ
$A_n + U_{vs} \rightarrow A_n + U_{vs}$	=	U_n
अ + ु → अु	=	उ

The combination of the independent form of the default vowel /a/ (in the Devanagari script, U+0905 DEVANAGARI LETTER A) with a dependent vowel sign may be viewed as an alternative spelling of the phonetic information normally represented by an isolated independent vowel form. However, these two representations should not be considered equivalent for the purposes of rendering. Higher-level text processes may choose to consider these alternative spellings equivalent in terms of information content, but such an equivalence is not stipulated by this standard.

Virama. Devanagari and other Indic scripts employ a sign known as the virama or vowel omission sign. In Hindi it is called *halant* and that term is used frequently in referring to the virama. A virama sign (for example, U+094D DEVANAGARI SIGN VIRAMA) nominally serves to cancel (or kill) the inherent vowel of the consonant to which it is applied. The virama functions as a combining character, with its shape varying from script to script. When a consonant has lost its inherent vowel by the application of virama, it is known as a *dead consonant*; in contrast, a *live consonant* is one that retains its inherent vowel or is written with an explicit dependent vowel sign. In the Unicode Standard, a dead consonant is defined as a sequence consisting of a consonant letter followed by a virama. The default rendering for a dead consonant is to position the virama as a combining mark bound to the consonant letterform.

For example, if C_n denotes the nominal form of consonant C, and C_d denotes the dead consonant form, then a dead consonant is encoded as shown in *Figure 9-2*.

Figure 9-2. Dead Consonants

$TA_n + VIRAMA_n \rightarrow TA_d$

त + ् → त्

Consonant Conjuncts. The Indic scripts are noted for a large number of consonant conjunct forms that serve as orthographic abbreviations (ligatures) of two or more adjacent letterforms. This abbreviation takes place only in the context of a *consonant cluster*. An orthographic consonant cluster is defined as a sequence of characters that represents one or more dead consonants (denoted C_d) followed by a normal, *live* consonant letter (denoted C_l) or an independent vowel ~~letter~~.

Under normal circumstances, a consonant cluster is depicted with a conjunct glyph if such a glyph is available in the current font(s). 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 (see *Figure 9-3*).

Figure 9-3. Conjunct Formations

(1) $GA_d + DHA_l \rightarrow GA_h + DHA_n$

ग् + ध → ग्ध

(2) $KA_d + KA_l \rightarrow K.KA_n$

क् + क → क्क

~~(3) $KA_d + SSHA_l \rightarrow K.SSHA_n$~~

~~क् + ष → क्ष~~

~~(4) $RA_d + RI_l \rightarrow RI_n + RA_{sup}$~~

~~र् + ऋ → र्ऋ~~

A number of types of conjunct formations appear in these examples: (1) a half-form of *GA* in its combination with the full form of *DHA*; (2) a vertical conjunct *K.KA*; ~~(3) a fully ligated conjunct *K.SSHA*, in which the components are no longer distinct; and (4) a rare conjunct formed with an independent vowel letter, in this case the vowel letter *RI* (also known as vocalic r). Note that in example (4) in *Figure 9-3*, the dead consonant RA_d is depicted with the nonspacing combining mark RA_{sup} (repha)~~.

A well-designed Indic script font may contain hundreds of conjunct glyphs, but they are not encoded as Unicode characters because they are the result of ligation of distinct letters. Indic script rendering software must be able to map appropriate combinations of characters in context to the appropriate conjunct glyphs in fonts.

~~When an independent vowel appears as the terminal element of a consonant cluster, as in example (4) in *Figure 9-3*, the independent vowel should not be depicted as a dependent vowel sign, but as an independent vowel letterform.~~

Explicit Virama. Normally a virama character serves to create dead consonants that are, in turn, combined with subsequent consonants to form conjuncts. This behavior usually results in a virama sign not being depicted visually. Occasionally, however, this default behavior is not desired when a dead consonant should be excluded from conjunct formation, in which case the virama sign is visibly rendered. To accomplish this goal, the Unicode Standard adopts the convention of placing the character U+200C ZERO WIDTH NON-JOINER immediately after the encoded dead consonant that is to be excluded from conjunct formation. In this case, the virama sign is always depicted as appropriate for the consonant to which it is attached.

~~For example, in *Figure 9-4*, the use of ZERO WIDTH NON-JOINER prevents the default formation of the conjunct form क्ष (K.SSHA$_n$).~~

Figure 9-4. Preventing Conjunct Forms

$$KA_d + ZWNJ + SSHA_l \rightarrow KA_d + SSHA_n$$

क् + $\boxed{\text{ZWNJ}}$ + ष → क्‌ष

Explicit Half-Consonants. When a dead consonant participates in forming a conjunct, the dead consonant form is often absorbed into the conjunct form, such that it is no longer distinctly visible. In other contexts, however, the dead consonant may remain visible as a *half-consonant form*. In general, a half-consonant form is distinguished from the nominal consonant form by the loss of its inherent vowel stem, a vertical stem appearing to the right side of the consonant form. In other cases, the vertical stem remains but some part of its right-side geometry is missing.

In certain cases, it is desirable to prevent a dead consonant from assuming full conjunct formation yet still not appear with an explicit virama. In these cases, the half-form of the consonant is used. To explicitly encode a half-consonant form, the Unicode Standard adopts the convention of placing the character U+200D ZERO WIDTH JOINER immediately after the encoded dead consonant. The ZERO WIDTH JOINER denotes a nonvisible letter that presents linking or cursive joining behavior on either side (that is, to the previous or following letter). Therefore, in the present context, the ZERO WIDTH JOINER may be considered to present a context to which a preceding dead consonant may join so as to create the half-form of the consonant.

For example, if C_h denotes the half-form glyph of consonant C, then a half-consonant form is encoded as shown in *Figure 9-5*.

Figure 9-5. Half-Consonants

$$KA_d + ZWJ + SSHA_l \rightarrow KA_h + SSHA_n$$

क् + $\boxed{\text{ZWJ}}$ + ष → क्‍ष

- ~~In the absence of the ZERO WIDTH JOINER, this sequence would normally produce the full conjunct form क्ष (K.SSHA$_n$).~~

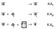
This encoding of half-consonant forms also applies in the absence of a base letterform. That is, this technique may also be used to encode independent half-forms, as shown in *Figure 9-6*.

Consonant Forms. In summary, each consonant may be encoded such that it denotes a live consonant, a dead consonant that may be absorbed into a conjunct, or the half-form of a dead consonant (*see Figure 9-7*).

Figure 9-6. Independent Half-Forms

GA_d + ZWJ → GA_h
ग् + ZWJ →

Figure 9-7. Consonant Forms



Rendering

Rules for Rendering. The following provides more formal and detailed rules for minimal rendering of Devanagari as part of a plain text sequence. It describes the mapping between Unicode characters and the glyphs in a Devanagari font. It also describes the combining and ordering of those glyphs.

These rules provide minimal requirements for legibly rendering interchanged Devanagari text. As with any script, a more complex procedure can add rendering characteristics, depending on the font and application.

It is important to emphasize that in a font that is capable of rendering Devanagari, the set of glyphs is greater than the number of Devanagari Unicode characters.

Notation. In the next set of rules, the following notation applies:

C_n	Nominal glyph form of consonant C as it appears in the code charts.
C_l	A live consonant, depicted identically to C_n.
C_d	Glyph depicting the dead consonant form of consonant C.
C_h	Glyph depicting the half-consonant form of consonant C.
L_n	Nominal glyph form of a conjunct ligature consisting of two or more component consonants. A conjunct ligature composed of two consonants X and Y is also denoted $X.Y_n$.
RA_{sup}	A nonspacing combining mark glyph form of the U+0930 DEVANAGARI LETTER RA positioned above or attached to the upper part of a base glyph form. This form is also known as *repha*.
RA_{sub}	A nonspacing combining mark glyph form of the U+0930 DEVANAGARI LETTER RA positioned below or attached to the lower part of a base glyph form.
V_{vs}	Glyph depicting the dependent vowel sign form of a vowel V.

$VIRAMA_n$ The nominal glyph form nonspacing combining mark depicting U+094D DEVANAGARI SIGN VIRAMA.

- A virama character is not always depicted; when it is depicted, it adopts this nonspacing mark form.

Dead Consonant Rule. The following rule logically precedes the application of any other rule to form a dead consonant. Once formed, a dead consonant may be subject to other rules described next.

R1 *When a consonant* C_n *precedes a* $VIRAMA_n$*, it is considered to be a dead consonant* C_d*. A consonant* C_n *that does not precede* $VIRAMA_n$ *is considered to be a live consonant* C_l*.*

$$TA_n + VIRAMA_n \rightarrow TA_d$$

त + ◌् → त्

Consonant* RA *Rules. The character U+0930 DEVANAGARI LETTER RA takes one of a number of visual forms depending on its context in a consonant cluster. By default, this letter is depicted with its nominal glyph form (as shown in the code charts). In some contexts, it is depicted using one of two nonspacing glyph forms that combine with a base letterform.

R2 *If the dead consonant* RA_d *precedes either a consonant or an independent vowel, then it is replaced by the superscript nonspacing mark* RA_{sup}*, which is positioned so that it applies to the logically subsequent element in the memory representation.*

$$RA_d + KA_l \rightarrow KA_l + RA_{sup} \qquad \text{Displayed Output}$$

र् + क → क + ◌ → र्क

$$RA^1_d + RA^2_d \rightarrow RA^2_d + RA^1_{sup}$$

र् + र् → र् + ◌ → र्र्

R3 *If the superscript mark* RA_{sup} *is to be applied to a dead consonant and that dead consonant is combined with another consonant to form a conjunct ligature, then the mark is positioned so that it applies to the conjunct ligature form as a whole.*

$$RA_d + JA_d + NYA_l \rightarrow J.NYA_n + RA_{sup} \qquad \text{Displayed Output}$$

र् + ज् + ञ → ज्ञ + ◌ → र्ज्ञ

R4 *If the superscript mark* RA_{sup} *is to be applied to a dead consonant that is subsequently replaced by its half-consonant form, then the mark is positioned so that it applies to the form that serves as the base of the consonant cluster.*

$$RA_d + GA_d + GHA_l \rightarrow GA_h + GHA_l + RA_{sup} \qquad \text{Displayed Output}$$

र् + ग् + घ → ग्‍ + घ + ◌ → र्ग्घ

R5 In conformance with the ISCII standard, the half-consonant form RRA_h *is represented as eyelash-*RA*. This form of* RA *is commonly used in writing Marathi.*

$$\mathrm{RRA}_n + \mathrm{VIRAMA}_n \rightarrow \mathrm{RRA}_h$$

ऱ + ◌् → ऱ्

R5a For compatibility with The Unicode Standard, Version 2.0, if the dead consonant RA_d *precedes* ZERO WIDTH JOINER, *then the half-consonant form* RA_h, *depicted as eyelash-*RA, *is used instead of* RA_{sup}.

$$\mathrm{RA}_d + \mathrm{ZWJ} \rightarrow \mathrm{RA}_h$$

र् + ZWJ → ऱ्

R6 Except for the dead consonant RA_d, *when a dead consonant* C_d *precedes the live consonant* RA_l, *then* C_d *is replaced with its nominal form* C_n, *and* RA *is replaced by the subscript nonspacing mark* RA_{sub}, *which is positioned so that it applies to* C_n.

$$\mathrm{TTHA}_d + \mathrm{RA}_l \rightarrow \mathrm{TTHA}_n + \mathrm{RA}_{sub} \quad \text{Displayed Output}$$

ठ् + र → ठ + ◌्र → ठ्र

R7 For certain consonants, the mark RA_{sub} *may graphically combine with the consonant to form a conjunct ligature form. These combinations, such as the one shown here, are further addressed by the ligature rules described shortly.*

$$\mathrm{PHA}_d + \mathrm{RA}_l \rightarrow \mathrm{PHA}_n + \mathrm{RA}_{sub} \quad \text{Displayed Output}$$

फ् + र → फ + ◌्र → फ्र

R8 If a dead consonant (other than RA_d*) precedes* RA_d, *then the substitution of* RA *for* RA_{sub} *is performed as described above; however, the* VIRAMA *that formed* RA_d *remains so as to form a dead consonant conjunct form.*

$$\mathrm{TA}_d + \mathrm{RA}_d \rightarrow \mathrm{TA}_n + \mathrm{RA}_{sub} + \mathrm{VIRAMA}_n \rightarrow \mathrm{T.RA}_d$$

त् + र् → त + ◌्र + ◌् → त्र्

A dead consonant conjunct form that contains an absorbed RA_d *may subsequently combine to form a multipart conjunct form.*

$$\mathrm{T.RA}_d + \mathrm{YA}_l \rightarrow \mathrm{T.R.YA}_n$$

त्र् + य → त्र्य

Modifier Mark Rules. In addition to vowel signs, three other types of combining marks may be applied to a component of an orthographic syllable or to the syllable as a whole: *nukta*, *bindus*, and *svaras*.

R9 *The nukta sign, which modifies a consonant form, is placed immediately after the consonant in the memory representation and is attached to that consonant in rendering. If the consonant represents a dead consonant, then* NUKTA *should precede* VIRAMA *in the memory representation.*

$$\mathrm{KA}_n + \mathrm{NUKTA}_n + \mathrm{VIRAMA}_n \rightarrow \mathrm{QA}_d$$

क + ◌़ + ◌् → क़्

R10 *The other modifying marks, bindus and svaras, apply to the orthographic syllable as a whole and should follow (in the memory representation) all other characters that constitute the syllable. In particular, the bindus should follow any vowel signs, and the svaras should come last. The relative placement of these marks is horizontal rather than vertical; the horizontal rendering order may vary according to typographic concerns.*

$$\mathrm{KA}_n + \mathrm{AA}_{vs} + \mathrm{CANDRABINDU}_n$$

क + ◌ा + ◌ँ → काँ

Ligature Rules. Subsequent to the application of the rules just described, a set of rules governing ligature formation apply. The precise application of these rules depends on the availability of glyphs in the current font(s) being used to display the text.

R11 *If a dead consonant immediately precedes another dead consonant or a live consonant, then the first dead consonant may join the subsequent element to form a two-part conjunct ligature form.*

$\mathrm{JA}_d + \mathrm{NYA}_l \rightarrow \mathrm{J.NYA}_n$ $\mathrm{TTA}_d + \mathrm{TTHA}_l \rightarrow \mathrm{TT.TTHA}_n$

ज् + ञ → ज्ञ ट् + ठ → ट्ठ

R12 *A conjunct ligature form can itself behave as a dead consonant and enter into further, more complex ligatures.*

$$\mathrm{SA}_d + \mathrm{TA}_d + \mathrm{RA}_n \rightarrow \mathrm{SA}_d + \mathrm{T.RA}_n \rightarrow \mathrm{S.T.RA}_n$$

स् + त् + र → स् + त्र → स्त्र

A conjunct ligature form can also produce a half-form.

~~$\mathrm{K.SSA}_d + \mathrm{YA}_l \rightarrow \mathrm{K.SS}_h + \mathrm{YA}_n$~~

क्ष् + य → क्ष्य

R13 If a nominal consonant or conjunct ligature form precedes RA_{sub} ~~*as a result of the application of rule R2*~~*, then the consonant or ligature form may join with* RA_{sub} *to form a multipart conjunct ligature* ~~*(see rule R2 for more information)*~~*.*

$KA_n + RA_{sub} \rightarrow K.RA_n$ $PHA_n + RA_{sub} \rightarrow PH.RA_n$

क + ्र → क्र फ + ्र → फ्र

R14 In some cases, other combining marks will also combine with a base consonant, either attaching at a nonstandard location or changing shape. In minimal rendering there are only two cases, RA_l *with* U_{vs} *or* UU_{vs} *.*

$RA_l + U_{vs} \rightarrow RU_n$ $RA_l + UU_{vs} \rightarrow RUU_n$

र + ु → रु र + ू → रू

Memory Representation and Rendering Order. The order for storage of plain text in Devanagari and all other Indic scripts generally follows phonetic order; that is, a CV syllable with a dependent vowel is always encoded as a consonant letter C followed by a vowel sign V in the memory representation. This order is employed by the ISCII standard and corresponds with both the phonetic and keying order of textual data ~~(see figure 9-8)~~.

Figure 9-8. Rendering Order

Character Order Glyph Order

$KA_n + I_{vs} \rightarrow I_{vs} + KA_n$

क + ि → कि

Because Devanagari 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_n denotes the nominal form of consonant C, and V_{vs} 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.

R15 When the dependent vowel I_{vs} *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. For example:*

$TA_d + RA_l + I_{vs} \rightarrow T.RA_n + I_{vs} \rightarrow I_{vs} + T.RA_n$

त् + र + ि → त्र + ि → त्रि

Sample Half-Forms. *Table 9-1* shows examples of half-consonant forms that are commonly used with the Devanagari script. These forms are glyphs, not characters. They may be encoded explicitly using ZERO WIDTH JOINER as shown; in normal conjunct formation, they may be used spontaneously to depict a dead consonant in combination with subsequent consonant forms.

Table 9-1. Sample Half-Forms

क	◌्	ZWJ	क्‍	न	◌्	ZWJ	न्‍
ख	◌्	ZWJ	ख्‍	प	◌्	ZWJ	प्‍
ग	◌्	ZWJ	ग्‍	फ	◌्	ZWJ	फ्‍
घ	◌्	ZWJ	घ्‍	ब	◌्	ZWJ	ब्‍
च	◌्	ZWJ	च्‍	भ	◌्	ZWJ	भ्‍
ज	◌्	ZWJ	ज्‍	म	◌्	ZWJ	म्‍
झ	◌्	ZWJ	झ्‍	य	◌्	ZWJ	य्‍
ञ	◌्	ZWJ	ञ्‍	ल	◌्	ZWJ	ल्‍
ण	◌्	ZWJ	ण्‍	व	◌्	ZWJ	व्‍
त	◌्	ZWJ	त्‍	श	◌्	ZWJ	श्‍
थ	◌्	ZWJ	थ्‍	ष	◌्	ZWJ	ष्‍
ध	◌्	ZWJ	ध्‍	स	◌्	ZWJ	स्‍

Sample Ligatures. *Table 9-2* shows examples of conjunct ligature forms that are commonly used with the Devanagari script. These forms are glyphs, not characters. Not every writing system that employs this script uses all of these forms; in particular, many of these forms are used only in writing Sanskrit texts. Furthermore, individual fonts may provide fewer or more ligature forms than are depicted here.

Table 9-2. Sample Ligatures

क	◌्	क	क्क	ट	◌्	ठ	ट्ठ
क	◌्	त	क्त	ठ	◌्	ठ	ठ्ठ
क	◌्	र	क्र	ड	◌्	ग	ड्ग
क	◌्	ष	क्ष	ड	◌्	ड	ड्ड
ङ	◌्	क	ङ्क	ड	◌्	ढ	ड्ढ
ङ	◌्	ख	ङ्ख	त	◌्	त	त्त
ङ	◌्	ग	ङ्ग	त	◌्	र	त्र
ङ	◌्	घ	ङ्घ	न	◌्	न	न्न
ञ	◌्	ज	ञ्ज	फ	◌्	र	फ्र
ज	◌्	ञ	ज्ञ	श	◌्	र	श्र
द	◌्	घ	द्घ	ह	◌्	म	ह्म
द	◌्	द	द्द	ह	◌्	य	ह्य
द	◌्	ध	द्ध	ह	◌्	ल	ह्ल

Table 9-2. Sample Ligatures (Continued)

द	्	ब	द्ब
द	्	भ	द्भ
द	्	म	द्म
द	्	य	द्य
द	्	व	द्व
ट	्	ट	ट्ट

ह	्	व	ह्व
ह		ृ	हृ
र		ु	रु
र		ू	रू
स	्	त्र	स्त्र

Sample Half-Ligature Forms. In addition to half-form glyphs of individual consonants, half-forms are also used to depict conjunct ligature forms. A sample of such forms is shown in *Table 9-3*. These forms are glyphs, not characters. They may be encoded explicitly using ZERO WIDTH JOINER as shown; in normal conjunct formation, they may be used spontaneously to depict a conjunct ligature in combination with subsequent consonant forms.

Table 9-3. Sample Half-Ligature Forms

क	्	ष	्	ZWJ	क्ष्
ज	्	ञ	्	ZWJ	ज्ञ्
त	्	त	्	ZWJ	त्त्
त	्	र	्	ZWJ	त्र्
श	्	र	्	ZWJ	श्र्

Combining Marks. Devanagari and other Indic scripts have a number of combining marks that could be considered diacritic. One class of these marks, known as bindus, is represented by U+0901 DEVANAGARI SIGN CANDRABINDU and U+0902 DEVANAGARI SIGN ANUSVARA. ~~These marks indicate nasalization or final nasal closure of a syllable.~~ U+093C DEVANAGARI SIGN NUKTA is a true diacritic. It is used to extend the basic set of consonant letters by modifying them (with a subscript dot in Devanagari) to create new letters. U+0951..U+0954 are a set of combining marks used in transcription of Sanskrit texts.

Digits. Each Indic script has a distinct set of digits appropriate to that script. These digits may or may not be used in ordinary text in that script. ~~European~~ digits have displaced the Indic script forms in modern usage in many of the scripts. Some Indic scripts—notably Tamil—lack a distinct digit for zero.

Punctuation and Symbols. U+0964 DEVANAGARI DANDA is similar to a full stop. Corresponding forms occur in many other Indic scripts. U+0965 DEVANAGARI DOUBLE DANDA marks the end of a verse in traditional texts. U+0970 DEVANAGARI ABBREVIATION SIGN appears after letters or combinations, for example [TX: 0921, 0949, 0970].

[[Need to insert glyphs above; assume just regular Unicode font? 2/4-02: 0921 (= 0xA1), 0949 (= 0xC9), 0970 (= 0Fx0?) check if these work in Unicode font)]]

Many modern languages written in the Devanagari script intersperse punctuation derived from the Latin script. Thus U+002C COMMA and U+002E FULL STOP are freely used in writing Hindi, and the *danda* is usually restricted to more traditional texts.

Encoding Structure. The Unicode Standard organizes the nine principal Indic scripts in blocks of 128 encoding points each. The first six columns in each script are isomorphic with the ISCII-1988 encoding, except that the last 11 positions (U+0955..U+095F in Devanagari, for example), which are unassigned or undefined in ISCII-1988, are used in the Unicode encoding.

The seventh column in each of these scripts, along with the last 11 positions in the sixth column, represent additional character assignments in the Unicode Standard that are matched across all nine scripts. For example, positions U+xx66..U+xx6F and U+xxE6..U+xxEF code the Indic script digits for each script.

The eighth column for each script is reserved for script-specific additions that do not correspond from one Indic script to the next.

Other Languages. Sindhi makes use of U+0974 DEVANAGARI LETTER SHORT YA. Several implosive consonants in Sindhi are realized as combinations with nukta and U+0952 DEVANAGARI STRESS SIGN ANUDATTA. Konkani makes use of additional sounds which can be made with combinations such as U+091A DEVANAGARI LETTER CA plus U+095C DEVANAGARI SIGN NUKTA and U+091F DEVANAGARI LETTER TTA plus U+0949 DEVANAGARI VOWEL SIGN CANDRA O.