



8. Unicode Standardization

8.1 Kannada Code Chart

	0C8	0C9	0CA	0CB	0CC	0CD	0CE	0CF
0		ಐ 0C90	ಠ 0CA0	ರ 0CB0	ೇ 0CC0		ಮ 0CE0	
1			ಡ 0CA1	ಱ 0CB1	ಱ 0CC1	ಱ 0CD1	ಱ 0CE1	
2	ಂ 0C82	ಒ 0C92	ಢ 0CA2	ಲ 0CB2	ೂ 0CC2	ಱ 0CD2		
3	ಃ 0C83	ಓ 0C93	ಣ 0CA3	ಳ 0CB3	ೈ 0CC3	ಱ 0CD3		
4		ಜ 0C94	ಱ 0CA4		ವ 0CC4	ಱ 0CD4		
5	ಱ 0C85	ಕ 0C95	ಢ 0CA5	ಱ 0CB5		ಱ 0CD5		ಱ 0CF5
6	ಱ 0C86	ಖ 0C96	ಢ 0CA6	ಱ 0CB6	ಱ 0CC6	ಱ 0CD6	ಱ 0CE6	
7	ಱ 0C87	ಗ 0C97	ಢ 0CA7	ಱ 0CB7	ಱ 0CC7		ಱ 0CE7	
8	ಱ 0C88	ಘ 0C98	ಱ 0CA8	ಱ 0CB8	ಱ 0CC8		ಱ 0CE8	
9	ಱ 0C89	ಙ 0C99		ಱ 0CB9			ಱ 0CE9	ಱ 0CF9
A	ಱ 0C8A	ಚ 0C9A	ಱ 0CAA	ಱ 0CBA	ಱ 0CCA		ಱ 0CEA	
B	ಱ 0C8B	ಛ 0C9B	ಱ 0CAB	ಱ 0CBB	ಱ 0CCB		ಱ 0CEB	
C	ಱ 0C8C	ಜ 0C9C	ಱ 0CAC	ಱ 0CBC	ಱ 0CCC		ಱ 0CEC	
D		ಝ 0C9D	ಱ 0CAD	ಱ 0CBD	ಱ 0CCD		ಱ 0CED	
E	ಱ 0C8E	ಞ 0C9E	ಱ 0CAE	ಱ 0CBE		ಱ 0CDE	ಱ 0CEE	
F	ಱ 0C8F	ಟ 0C9F	ಱ 0CAF	ಱ 0CBF			ಱ 0CEF	



8.1.1 Kannada Code Chart Details

Code Character Description
Point

Various signs

0C82 ೀ KANNADA SIGN ANUSVARA

0C83 ೆ KANNADA SIGN VISARGA

Independent vowels

0C85 ಅ KANNADA LETTER A

0C86 ಆ KANNADA LETTER AA

0C87 ಇ KANNADA LETTER I

0C88 ಈ KANNADA LETTER II

0C89 ಉ KANNADA LETTER U

0C8A ಊ KANNADA LETTER UU

0C8B ಋ KANNADA LETTER
VOCALIC R

0C8C ೠ KANNADA LETTER
VOCALIC L
• Not in present use

0C8D <reserved>

0C8E ಎ KANNADA LETTER E

0C8F ಏ KANNADA LETTER EE

0C90 ಐ KANNADA LETTER AI

0C91 <reserved>

0C92 ಒ KANNADA LETTER O

0C93 ಓ KANNADA LETTER OO

0C94 ಔ KANNADA LETTER AU

Consonants

0C95 ಕ KANNADA LETTER KA

0C96 ಖ KANNADA LETTER KHA

0C97 ಗ KANNADA LETTER GA

0C98 ಘ KANNADA LETTER GHA

0C99 ಙ KANNADA LETTER NGA

0C9A ಚ KANNADA LETTER CA

0C9B ಛ KANNADA LETTER CHA

0C9C ಜ KANNADA LETTER JA

0C9D ಝ KANNADA LETTER JHA

0C9E ಞ KANNADA LETTER NYA

0C9F ಟ KANNADA LETTER TTA

0CA0 ಠ KANNADA LETTER TTHA

0CA1 ಡ KANNADA LETTER DDA

0CA2 ಢ KANNADA LETTER DDHA

0CA3 ಣ KANNADA LETTER NNA

0CA4 ತ KANNADA LETTER TA

0CA5 ಥ KANNADA LETTER THA

0CA6 ದ KANNADA LETTER DA

0CA7 ಧ KANNADA LETTER DHA

0CA8 ನ KANNADA LETTER NA

0CA9 <reserved>

0CAA ಪ KANNADA LETTER PA

0CAB ಫ KANNADA LETTER PHA

0CAC ಬ KANNADA LETTER BA

0CAD ಭ KANNADA LETTER BHA

0CAE ಮ KANNADA LETTER MA

0CAF ಯ KANNADA LETTER YA

0CB0 ರ KANNADA LETTER RA

0CB1 ಱ KANNADA LETTER RRA

0CB2 ಲ KANNADA LETTER LA

0CB3 ಳ KANNADA LETTER LLA

0CB4 <reserved>

0CB5 ವ KANNADA LETTER VA

0CB6 ಷ KANNADA LETTER SHA

0CB7 ಷ KANNADA LETTER SSA

0CB8 ಸ KANNADA LETTER SA

0CB9 ಹ KANNADA LETTER HA

0CBA ೀ KANNADA INVISIBLE

LETTER

0CBB ೆ KANNADA VOWEL SIGN A

0CBC ೆ KANNADA SIGN NUKTA

0CBD ೆ KANANADASIGN AVAGRAHA



Dependent vowel signs

0CBE	ೠ	KANNADA VOWEL SIGN AA
0CBF	ೡ	KANNADA VOWEL SIGN I
0CC0	ೢ	KANNADA VOWEL SIGN II
0CC1	ೣ	KANNADA VOWEL SIGN U
0CC2	೤	KANNADA VOWEL SIGN UU
0CC3	೶	KANNADA VOWEL SIGN VOCALIC R
0CC4	೷	KANNADA VOWEL SIGN VOCALIC RR
0CC5		<reserved>
0CC6	೹	KANNADA VOWEL SIGN E
0CC7	೺	KANNADA VOWEL SIGN EE
0CC8	೻	KANNADA VOWEL SIGN AI
0CC9		<reserved>
0CCA	೽	KANNADA VOWEL SIGN O
0CCB	೾	KANNADA VOWEL SIGN OO
0CCC	೿	KANNADA VOWEL SIGN AU

Various signs

0CCD	೿	KANNADA SIGN HALANT
0CCE		<reserved>
0CCF		<reserved>
0CD0		<reserved>
0CD1	೿	KANNADA DIACRITIC SIGN UDATTA •Used above any character
0CD2	೿	KANNADA DIACRITIC SIGN ANUDATTA •Used below any character
0CD3	೿	KANNADA DIACRITIC SIGN GURU-GRAVE •Used above any character
0CD4	೿	KANNADA DIACRITIC SIGN LAGHU-ACUTE •Used above any character

0CD5	೿	KANNADA LENGTH MARK
0CD6	೿	KANNADA AI LENGTH MARK

Additional consonants

0CDE	೿	KANNADA LETTER LLLA
------	---	---------------------

Generic additions

0CE0	೿	KANNADA LETTER VOCALIC RR
------	---	---------------------------

0CE1	೿	KANNADA LETTER VOCALIC LL
------	---	---------------------------

•Not in present use

Digits

0CE6	೿	KANNADA DIGIT ZERO
0CE7	೿	KANNADA DIGIT ONE
0CE8	೿	KANNADA DIGIT TWO
0CE9	೿	KANNADA DIGIT THREE
0CEA	೿	KANNADA DIGIT FOUR
0CEB	೿	KANNADA DIGIT FIVE
0CEC	೿	KANNADA DIGIT SIX
0CED	೿	KANNADA DIGIT SEVEN
0CEE	೿	KANNADA DIGIT EIGHT
0CEF	೿	KANNADA DIGIT NINE
0CF5	೿	KANNADA SIGN REPH
0CF9	೿	KANNADA DIACRITIC SIGN DEERGHASWARITHA

•Used above any character



8.1.2 Kannada General Information & Description

Introduction

The Kannada script is a South Indian script. It is used to write Kannada language of Karnataka State in India. This is also used in many parts of Tamil Nadu, Kerala, Andhra Pradesh and Maharashtra. In addition, the Kannada script is also used to write Tulu, Konkani and Kodava languages. Kannada along with other Indian language scripts shares a large number of structural features. The Kannada block of Unicode Standard (0C80 to 0CFF) is based on ISCII-1988 (Indian Standard Code for Information Interchange). The Unicode Standard (version 3) encodes Kannada characters in the same relative positions as those coded in the ISCII-1988 standard.

The Writing system that employs Kannada script constitutes a cross between syllabic writing systems and phonemic writing systems (alphabets). The effective unit of writing Kannada 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 Kannada script. These consist of distinct character types: Consonant letters, independent vowels and the corresponding dependent vowel signs. In a text sequence, these characters are stored in logical phonetic order.

Rendering Kannada Characters

Kannada characters can combine or change shape depending on their context. A character's appearance is affected by its ordering with respect to other characters and the application or system environment. This variation can cause the appearance of Kannada characters to be different from nominal glyphs.

Vowels (Swaras)

Independent vowel letters

The independent vowels (*Swaras*) in Kannada 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, which start with a vowel. The Unicode character encoding for Kannada uses a distinct set of naming conventions for some mid vowels of the fourteen vowels in Kannada. Of these fourteen vowels, twelve vowels have been divided into six sets, each set consisting of a short vowel (*Hrasva Swara*), followed by a corresponding long vowel (*Deergha Swara*). These are two types of *Swaras* depending on the time used to pronounce them.

Hrasva Swara is a freely existing independent vowel which can be pronounced in a single *matra* time (matra kala) whereas a *Deergha Swara* is the vowel which can be pronounced in two *matra*.time. The six sets of the swaras are :

ಅ, ಆ (0C85 , 0C86)
ಇ, ಈ (0C87 , 0C88)
ಉ, ಊ (0C89 , 0C8A)
ಋ, ೠ (0C8B , 0CE0)
ಎ, ಏ (0C8E , 0C8F)
ಒ, ಓ (0C92 , 0C93)

Of these, the vowel ಋ(0CE0) is not in present use.

The two Deergha swaras ಐ(0C90) and ಔ(0C94) have no Hrasva swara counterparts.

Further, the so-called swaras with code values 0C8C and 0CE1 are not used in Kannada and are not required for Kannada.

Dependent vowel signs (Matras)

The dependent vowel signs serve as the common manner of writing non-inherent vowels and are generally referred to as *Swara Chinhas* in Kannada or *Matras* in Samskrit. The dependent vowel signs do not appear stand-alone; rather, they are visibly depicted in combination with a base-letter form (generally a consonant). A single consonant or a consonant cluster may have a dependent vowel sign 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 sign in a syllable overrides the inherent vowel ಅ(0C85) of a single consonant letter.

There are several variations with which the dependent vowels are applied to the base letterforms. Most of them appear as non-spacing dependent vowel signs when applied to base letterforms; above or to the right side of a consonant letter or a consonant cluster. The following are the exceptions and variations for the above rule:



- A. The two dependent vowel signs ೀ(0CC3) & ೆ(0CC4) appear one level below and to the right of the consonant or the consonant cluster separated by a small white space.
- B. Each of the five dependent vowels signs ೀ (U+0CC0), ೆ(0CC7), ೆ̣(0CC8), ೆ̣̣(0CCA) & ೆ̣̣̣(0CCB) are depicted by two or three glyph components (two part or three part vowel signs) with one component appearing with a space to the right of the consonant or the consonant cluster
- i) In the case of three of the above-mentioned two/three-part dependent vowels ೆ (0CC0), ೆ (0CC7) and ೆ̣̣̣(0CCB), the non-spacing component(s) of each of them is(are) the same as the vowel sign(s) of the corresponding preceding short vowels. The spacing component for each of these dependent vowels is the same length mark ೆ (0CD5) given in Unicode version 3. The logic for this is that these dependent vowels are nothing but the long forms (independent and phonetically distinct) of the preceding short vowels.
- ii) The first component of the dependent vowel ೆ̣̣̣ (0CC8) mentioned above is the same as the dependent vowel ೆ̣̣ (0CC6) and the second component is same as ೆ̣̣ (0CD6). These are defined independently in Unicode version 3. The second part appears slightly below and to the right of the consonant or the consonant clusters.
- C. In view of this, it is important to note that the two glyphs (the length mark ೆ and the second component of ೆ̣̣̣ i.e. ೆ̣̣) represented with the codes at 0CD5 and 0CD6 in Unicode version 3 have no independent existence and do not play any part as independent codes in the collation algorithm.
- D. Unlike Devanagari, the Kannada script does not have any character with a left-side dependent vowel sign.
- E. A one-to-one correspondence exists between independent vowels and dependent vowel signs.

Consonant letters (Vyanjanas)

Each of the 36 consonant letters in Kannada (enumerated with codes 0C95 to 0CB9 and 0CDE) represents a single consonantal sound but also has

the peculiarity of having an inherent vowel, generally the short vowel ಅ (/a/ 0C85).

Thus the Kannada letter at 0C95 represents not just ಕ (K) but ಕ (KA) with the inherent vowel ಅ(0C85). In the presence of a dependent vowel, however, this inherent vowel associated with a consonant letter is overridden by the dependent vowel. The consonants ಂ (0CB1) and ಂ(0CDE) sound similar to ಠ(0CB0) and ಱ(0CB3) respectively. These two appear in ancient Kannada texts but are not in present use. With this, consonants in modern Kannada are 34 in number (without ಂ and ಂ). These are classified as *Vargeeya Vyanjanas*(0C95 to 0CAE) and *Avargeeya Vyanjanas* (0CAF, 0CB0 and 0CB2 to 0CB9).

Vargeeya Vyanjanas : The five sets of *Vargeeya Vyanjanas* are

ಕ	ಖ	ಗ	ಘ	ಙ
0C95	0C96	0C97	0C98	0C99
ಚ	ಛ	ಜ	ಝ	ಞ
0C9A	0C9B	0C9C	0C9D	0C9E
ಟ	ಠ	ಡ	ಢ	ಣ
0C9F	0CA0	0CA1	0CA2	0CA3
ತ	ಥ	ದ	ಧ	ನ
0CA4	0CA5	0CA6	0CA7	0CA8
ಪ	ಫ	ಬ	ಭ	ಮ
0CAA	0CAB	0CAC	0CAD	0CAE

Avargeeya Vyanjanas : The nine Avargeeya Vyanjanas (enumerated in the acceptable sorting order) are:

ಯ	ರ	ಲ	ವ	ಶ	ಷ	ಸ	ಹ	ಳ
0CAF	0CB0	0CB2	0CB5	0CB6	0CB7	0CB8	0CB9	0CB3

Halant

Like Devanagari, Kannada script also employs a sign known as *halant* or vowel omission sign. A halant sign (ೆ̣̣̣̣, 0CCD) nominally serves to cancel (or kill) the inherent vowel of the consonant to which it is applied.

The *halant* functions as a combining character. When a consonant has lost its inherent vowel by the application of *halant*, it is known as a dead consonant. The dead consonants are the presentation forms used to depict the consonants without an inherent vowel. Their rendered forms in Kannada resemble the full consonant with the vertical stem replaced by the *halant* sign, which marks a character core. The stem glyph (ೆ̣̣̣̣ at 0CBB)



is graphically and historically related to the sign denoting the inherent /a/ (ॐ) vowel (0C85). In contrast, a live consonant is a consonant that retains its inherent vowel or is written with an explicit dependent vowel sign. The dead consonant is defined as a sequence consisting of a consonant letter followed by a *halant*. The default rendering for a dead consonant is to position the *halant* as a combining mark bound to the consonant letter form.

Avagraha (ॐ)

A spacing mark ॐ, called avagraha sign is used while rendering Samskrit texts. This is located at 0C8D.

Encoding order

The traditional Kannada 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 (Indian Script Code for Information Interchange) & the Unicode standard both observe this traditional order.

Consonant conjuncts (Samyuktaksharas)

Like any other Indian script, Kannada is also noted for a large number of consonant conjunct forms that serve as orthographic abbreviations (ligatures) of two or more adjacent forms. This abbreviation takes place only in the context of a consonant cluster. An orthographic consonant cluster is defined as a sequence of characters that represent one or more dead consonants (denoted by C_d) followed by a normal live consonant (denoted by C_l).

Corresponding to each Kannada consonant, there exists a separate and unique glyph, which is specially used to represent the corresponding consonant in a consonant cluster. Most of these conjunct consonant glyphs resemble their original consonant forms (many without the implicit vowel sign, wherever applicable).

In Kannada, there is only one type of conjunct formation (consonant cluster) and it is depicted as follows:

- ▶ The first consonant of the consonant cluster is rendered with the implicit or a different dependant vowel appearing as the terminal element of the consonant cluster.

- ▶ The remaining consonants (consonants in between the first consonant and the terminal vowel element) appear in conjunct consonant glyph forms in the phonetic order. They are generally depicted directly below or sometimes below but to the right of the first consonant.

Thus, the systematically designed Kannada script font contains the conjunct glyph components, but they are not encoded as Unicode characters, because they are the resultant of ligation of distinct letters. Kannada script rendering software must be able to map appropriate combinations of characters in context to the appropriate conjunct glyphs in fonts.

Invisible consonant INV

There is a need to have a consonant, which provides an invisible base for the display of dependant vowels without any consonant base. This can be the Unicode Standard Zero Width Non-Joiner at 200C. This can also be used to provide proper collation of the words containing dead consonants.

Explicit *Halant*

Normally, a halant character serves to create dead consonants, which, in turn, combine with subsequent consonants in order to form conjuncts. This behavior usually results in a *halant* 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 *halant* sign is visibly rendered.

In order to accomplish this, the Unicode Standard character 200C (Zero Width Non-Joiner) is introduced immediately after the encoded dead consonant that is to be excluded from conjunct formation.

For example, the use of Zero Width Non-Joiner prevents the default formation of the conjunct form ಧ್ಢ, resulting in ಧ್ಢ.

The Kannada script adopts the convention of depicting the character (in this case the halant sign) as appropriate for the consonant to which it is attached.

In summary, each Kannada consonant may be encoded such that it denotes a live consonant, a dead consonant or a conjunct consonant glyph.



Memory Representations and Rendering Order

Notation

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

C_n	Nominal glyph form of a 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 a consonant C.
C_h	Glyph depicting the half-consonant form of a 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 by $X.Y_n$.
RA_{sub}	A non-spacing combining mark glyph form positioned below the base glyph form.
V_{vs}	Glyph depicting the dependent vowel sign form of a vowel V.
$HALANT_n$	The nominal glyph form non-spacing combining mark depicting 0CCD Kannada sign Halant.

A halant character is not always depicted; when it is depicted, it adopts this non-spacing mark form.

Memory Representations and Rendering Order

The order for storage of plain text in Kannada generally follows the phonetic order, that is, a CV syllable with a dependant 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 phonetic and keying order of textual data. Unlike Devanagari and some other Indian Scripts, all the dependent vowels in Kannada are depicted to the right of their consonant letters. Hence there is no need to reorder the elements in mapping from the logical (character) store to the presentation (glyph) rendering and vice versa.

Rule R1 : Whenever a consonant is followed by a vowel, then the corresponding vowel sign attaches to the consonant suitably.

Character order				Glyph order
$KA_n + U$	\rightarrow	$KA_n + U_{vs}$		
ಕ + ಉ	\rightarrow	ಕು		

Further, Kannada script does not allow half-consonants, ligatures and half ligature forms. The following provides more formal and complete rules for minimal rendering of Kannada as part of a plain text sequence. It describes the mapping between Unicode characters and the glyphs in a Kannada font. It also describes the combining and ordering of those glyphs.

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

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.

Rule R2 : When a consonant C_n precedes a $HALANT_n$, it is considered to be a dead consonant C_d . A consonant C_n that does not precede $HALANT_n$ is considered to be live consonant C_l .

$KA_n + HALANT_n$	\rightarrow	KA_d
ಕ + ಳ	\rightarrow	ಕಳ

Consonant cluster (conjunct) rendering

As already explained in section 8, the conjunct formation (consonant cluster) with two or more consonants and a terminal vowel is as follows:

- The first consonant of the consonant cluster is rendered with the terminal vowel.
- The remaining consonants (in between the first consonant and the terminal vowel) are rendered in conjunct consonant glyph forms in the phonetic order.

Rule R3 : Example1:

$KA_d + KA_n$	\rightarrow	KA_h
ಕಳ + ಕ	\rightarrow	ಕಳಕ

(conjunct consonant glyph of f to ಳ)

Example 2 :

$SA_d + TH_d + RA_d + I_{vs}$	\rightarrow	$SAI_{vs} + TH_h + RA_h$
ಸಳ + ತಳ + ಳ + ಿ	\rightarrow	ಸಿಲೆ

($\overset{\curvearrowright}{\text{ಲ}}$ and $\overset{\curvearrowright}{\text{ಲ}}$ are the conjunct consonant glyphs of ತಳ and ಳ)



Consonant Clusters with two different display forms : Consonant RA Rules

Whenever a consonant cluster of two or more consonants is formed with the Kannada consonant letter RA (ರ, 0CB0) as the first component of the consonant cluster, the component of this letter RA is depicted with two different presentation forms: one as the initial and the other as the final display element of the consonant cluster as detailed below.

Consonant clusters with RA as the first consonant: general method of rendering

Rule R4 : As explained before, the character ರ is rendered with the terminal vowel (implicit or dependent) and the in-between consonants are rendered below and/or to the right of ರ, in conjunct consonant glyph forms (ರೃ, ರೞ etc.).

Example 1:

$RA_d + KA_1 \rightarrow RA_1 + KA_h$
ರ + ಕ → ರೃ

Example 2 :

$RA_d + MA_1 + U_{vs} \rightarrow RA_n + MA_h + U_{vs}$
ರ + ಮ + ಁ → ರ್ಮು

Example 3 :

$RA_d + TA_d + YA_n \rightarrow RA_n + TA_h + YA_h$
ರ + ತ + ಯ → ರ್ತಯ

Consonant clusters with RA as the first consonant: Alternate method of rendering

Rule R5 : In the alternate representation method also, the above procedure is followed assuming ರ is absent (which means that the conjunct formation starts from the second consonant) to obtain the consonant cluster (conjunct). This is followed by another distinct glyph ಼ for ರ and this new glyph is depicted to the extreme right of the conjunct formed above. As per this representation, the conjuncts ರೃ, ರ್ಮು and ರ್ತಯ rendered in examples 1, 2 and 3 above are rendered as ಕಱ, ಮುಱ and ತೃಱ. The corresponding rule is as follows:

Example 1:

$RA_d + KA_1 \rightarrow KA_1 + Arkavottu$
ರ + ಕ → ಕಱ

Example 2:

$RA_d + MA_1 + U_{vs} \rightarrow MA_n + U_{vs} + Arkavottu$
ರ + ಮ + ಁ → ಮುಱ

Example 3:

$RA_d + TA_d + YA_n \rightarrow TA_n + YA_h + Arkavottu$
ರ + ತ + ಯ → ತೃಱ

Exception for the alternate method

Rule R6 : The exception for the rule R4 is that, whenever a conjunct is formed with both the first and second consonants as ರ (RA) (ie. a consonant conjunct using ರ with ರ itself, the rule R5 will not hold good. Instead, the general method of consonant conjunct formation is used (Rule R4). This means the conjunct consonant glyph ಱ of ರ is rendered.

$RA_d + RA_1 + O \rightarrow RA_n + RA_h + O_{vs}$
ರ + ರ + ಓ → ರ್ಱೋ

Nukta- Modifier Mark Rules

In addition to the vowel signs, one more type of combining mark may be applied to a component of an orthographic syllable or the syllable as a whole. The *NUKTA* sign, which modifies a consonant form, is placed immediately after the consonant (after the terminating vowel in case of a dependent vowel appearing after the consonant) in the memory representation and is attached to that consonant in rendering. If the consonant represents a dead consonant, then the *nukta* should precede halant in the memory representation. The *nukta* is represented by a double-dot mark placed at the location 0CBC. Two such modified consonants used in Kannada are ಙ (Pronounced as ZA) and ಞ (Pronounced as FĀ).

Diacritics

Diacritics are the principle class of non-spacing combining characters used with the Indian scripts. Diacritic is defined very broadly to include accents as well as other non-spacing marks. Kannada has a number of combining marks that could be considered diacritic. A set of five combining marks Udattha (ˆ above the character), Anudattha (˘ below the character), Guru (ˉ above the character), Laghu (˘ above the character) and Deergha Swaritha (ˆ above the character) located at 0CD1, 0CD2, 0CD3, 0CD4 and 0CF9 respectively. These are used in the transcription of Sanskrit texts (where ever needed) and for Kannada grammatical notations.

Digits

As in many Indian languages, Kannada also has a distinct set of appropriate digits. These are being used widely in ordinary texts, Government and public places. These are enumerated with code numbers 0CE6 to 0CEF.



Part-2

Sorting issues in Kannada

The sorting sequence for Kannada in Unicode is as per the collation chart enclosed with this document. However, the following are some important issues, which have to be addressed separately for proper sorting of data in Kannada.

ISCII – 91 provides direct sorting through its codes. It is the natural sorting method just based on code values. There are no special algorithms for language specific issues for sorting the data. This results in non-conventional sorting in some specific cases. The scholars in Kannada have specified the sorting standards in Kannada. These standards are being followed in all dictionaries and other documents in Kannada. With this in view, the following four special cases have been identified.

Sorting of *Nukta* characters

The modifying mark or *Nukta* located at 0CBC and included in the collation table is enough to take care of the sorting issues of characters ಙ (modified ಙ) and ಞ (modified ಞ). It also takes care of any other consonant, which may be modified using *Nukta*.

Sorting the data records containing *anuswara* and *visarga*

Sorting a data set containing words **terminating with *anuswara*, *Visarga* together with other words.** In such cases, words without terminating dependent vowels are placed in wrong positions.

- Sorting sequence as per the Unicode is according to the specified standards if the *anuswara* and *visarga* appear within a word.

Sorting of words with dead consonants

- **Sorting of words terminating with dead consonants**

Sorting in this case also violates the sorting rules of Kannada. The Unicode sorting places the word

terminating with the dead consonant at the end of the list. The following list compares the sorting of a sample data using Unicode table and the acceptable sorting for this case.

Sorted data as per Unicode	Acceptable sorting
ರಾಕ	ರಾಕ್
ರಾಕ್	ರಾಕ
ರಾಗ	ರಾಗ್
ರಾಗೀ	ರಾಗ
ರಾಗ್	ರಾಗೀ

- **Dead consonants within words**

Proper sorting of data with such words can be achieved by using the invisible zero width consonant just after the dead consonant.

To circumvent unacceptable situations mentioned in sections 2.2 and 2.3 above, the Unicode Standard character 200C (Zero Width Non-Joiner) can be used appropriately in the pre-processor and collation algorithms.

Sorting of Conjuncts having two different display forms

Two such conjuncts are rendered in Kannada at present.

- **Conjuncts with ರ (0CB0) as the first consonant**

This has been explained at an earlier section as **Consonant Ra rules.**

Words containing both the display forms of the same consonant cluster with ರ (0CB0) as the first consonant of the cluster has to be sorted as follows. Even though the display rendering are different, both are identical in all respects. It is therefore natural that they should appear at consecutive positions. Even though a separate glyph and a corresponding glyph code are present in the display/storage codes, such an arrangement in Unicode will not render for proper sorting.

The only alternative is to represent both the display forms by the same set of codes with a



distinguishing code (0CF5) within the string for the second display form. In Unicode form, the distinguishing code value within the string of the consonant cluster for the second display form is to be considered as ignorable for the purpose of sorting (Ref. Implementation Guidelines, Section 5.17 of Unicode Standard Version 3 document). This can be achieved through preprocessing software, with specific functions to generate proper glyph codes, storage codes, and the Unicode at different levels. Such a situation-specific code representation guarantees proper sorting of data containing consonant clusters with two different display forms by ignoring the code 0CF5 for ϵ . This condition has to be incorporated at the appropriate place in the sorting algorithm.

- The second case of rendering a same character in two different display forms is the dead consonant ಞ . It is also written in a second form as ಞ . Sorting issue in regard to this case is also dealt with the same way as in the previous case.

The Zero Width Non-Joiner at 200C cannot be used instead of ϵ (0CF5), as the same sequence of characters appear both with Zero Width Non-joiner and with ϵ , the two sequences representing two different syllables (conjuncts).

Sorting of Diacritic characters

Diacritic characters formed using symbols located at 0CD1, 0CD2, 0CD3 0CD4 and 0CF9 to render accents to consonants, are considered to be equivalent to the corresponding consonants for sorting purposes and hence the above procedure can be adopted in such cases also.

Conclusion

The sorting issues mentioned above may have multiple solutions. Similar issues might have been solved by different methods in respect of other Indian languages. Hence, it is desirable to evolve uniform procedures for issues common to all the

Indian languages. However, solutions for sorting problems mentioned here with respect Kannada have been obtained by considering all the consonants from 0C95 to 0CB9 and the consonant 0CDE when they appear independently in a data field as pure consonants (i.e. as two part coded [Ex: 0C95 \equiv (0C95, 0CBB)]). The sorting of a data field is achieved by the indexing method. All these can be elaborated to give the actual algorithms and flow charts, if need be.

Acknowledgements

Acknowledgements are due from Directorate of Information Technology, Govt. of Karnataka, to the following persons who have taken the responsibility in arriving at the Unicode standard and prepared this document.

- Mr. C V Srinatha Sastry, Assistant Director, National Aerospace Laboratories, Bangalore 560 017, General Secretary, Kannada Ganaka Parishath, Bangalore 560 019 and Member, Technical Advisory Committee on Standardisation and Usage of Kannada on Computers, Government of Karnataka.
- Dr. U B Pavanaja, CEO, Visva Kannada Softec, Bangalore and Member, Technical Advisory Committee on Standardisation and Usage of Kannada on Computers, Government of Karnataka and Member, Kannada Ganaka Parishath, Bangalore.
- Mr. G N Narasimha Murthy, Secretary, Kannada Ganaka Parishath, Bangalore and Manager, State Bank of India.
- Prof. G Venkatasubbiah, Former President, Kannada Sahithya Parishath and Former Professor, Vijaya College, Bangalore.
- Prof M H Krishnaiah, Former Professor, Bangalore University
- Prof. Narahalli Balasubrahmanya, Professor, Bangalore University.



Column 1	Column 2	Column 3	Column 4	Column 5
0C82	0CCD	0C96	0CA6	0CB9
೦	ಠ	ಖ	ದ	ಹ
0C83	0CBB	0C97	0CA7	0CB3
:	ಠ	ಗ	ಧ	ಳ
0C85	0CBE	0C98	0CA8	0CB4
ಅ	ಠ	ಘ	ನ	ಱ
0C86	0CBF	0C99	0CAA	0CBC
ಆ	ಠ	ಜ	ಪ	.
0C87	0CC0	0C9A	0CAB	0C88
ಇ	ಠ	ಚ	ಘ	ಈ
0CC1	0C9B	0CAC	0C89	0CC2
ಁ	ಠ	ಬ	ಉ	ಊ
0C9C	0CAD	0C8A	0CC3	0C9D
ಜ	ಭ	ಊ	ಠ	ಠ
0CAE	0C8B	0CC4	0C9E	0CAF
ಮ	ಋ	ಠ	ಠ	ಯ
0CE0	0CC6	0C9F	0CB0	0C8E
ಋ	ಠ	ಠ	ಠ	ಠ
0CC7	0CA0	0CB1	0C8F	0CC8
ಠ	ಠ	ಠ	ಠ	ಠ
0CA1	0CB2	0C90	0CCA	0CA2
ಠ	ಠ	ಠ	ಠ	ಠ
0CB5	0C92	0CCB	0CA3	0CB6
ಠ	ಠ	ಠ	ಠ	ಠ
0C93	0CCC	0CA4	0CB7	0C94
ಠ	ಠ	ಠ	ಠ	ಠ
0C95	0CA5	0CB8		
ಠ	ಠ	ಠ		

Collating sequence of Kannada Unicode Characters.

(Courtesy : Shri C. V. Srinath Sastry
 Directorate of Information Technology
 Government of Karnataka,
 Bangalore (Karnataka)
 Phone: 080-5279611(O), 6645865(R)
 E-mail : sastry@ead.cmmacs.ernet.in)

8.1.3 Typical Colloquial Sentences in Kannada

GREETING

- ▶ Hello
ನಮಸ್ಕಾರ
Namaskara
ನಮಸ್ಕಾರ
- ▶ Good Morning
ಶುಭ ಮುಂಜಾನೆ, ಶುಭೋದಯ
Shubha munjaane
ಶುಭ ಮುಂಜಾನೆ
- ▶ Good Afternoon
ಶುಭ ಮಧ್ಯಾಹ್ನ
Shubha Madhyanha
ಶುಭ ಮಧ್ಯಾಹ್ನ
- ▶ Good Night
ಶುಭ ರಾತ್ರಿ
Shubha Raathri
ಶುಭ ರಾತ್ರಿ
- ▶ Good Bye
ಶುಭ ಹಾರೈಕೆಯ ಮಾತು(NA)ಗುಡ್ ಬೈ
Shubha Haaraikeya maathu-Good bye
ಶುಭ ಹರೈಕೆಯ ಮಾತು-ಗುಡ್ ಬೈ
- ▶ Thanks
ಧನ್ಯವಾದಗಳು
Dhanyavaadhagalu
ಧನ್ಯವಾದಗಲು
- ▶ How are you
ನೀವು ಹೇಗಿದ್ದೀರಿ/ಹೇಗಿದ್ದೀಯ
neevu Hegiddeeri/Hegidhdheeya
ನೀವು ಹೇಗಿದ್ದೀರಿ/ಹೇಗಿದ್ದೀಯ
- ▶ I am fine thank you
ನಾನು ಚೆನ್ನಾಗಿದ್ದೇನೆ. ಧನ್ಯವಾದಗಳು
Naanuchennagiddene.Dhanyavaadhagalu
ನಾನು ಚೆನ್ನಾಗಿದ್ದೇನೆ. ಧನ್ಯವಾದಗಲು
- ▶ Sorry
ಪಶ್ಚಾತ್ತಾಪಪಡು/ಕ್ಷಮಿಸಿ
pashchyaathaapapadu/kshamisi
ಪಶ್ಚಾತ್ತಾಪ ಪಡು

GREETING

- ▶ It is cold
ಚಳಿ ಇದೆ
chali idhe
ಚಳಿ ಇದೆ



► It is cool outside

ಹೊರಗಡೆ ತಂಪಾಗಿದೆ.

Horagade Thampaagidhe

ಹೊರಗಡೆ ತಂಪಾಗಿದೆ

► It is hot

ಬಿಸಿಲು/ಬೇಸಗೆ/ಧಗೆ ಇದೆ

Bisilu/Besige/Thampu

ಬಿಸಿಲು/ ಬೇಸಗೆ/ಧಗೆ

► It is raining

ಮಳೆ ಬರುತ್ತಿದೆ.

Male Baruthidhe

ಮಳೆಬರುತ್ತಿದೆ

GENERAL

► What is your name?

ನಿನ್ನ/ನಿಮ್ಮ ಹೆಸರೇನು?

Ninna/Nimma Hesarenu?

ನಿನ್ನ/ನಿಮ್ಮ ಹೆಸರೇನು?

► My name is Ranjan

ನನ್ನ ಹೆಸರು ರಂಜನ್

Nanna Hesaru Ranjan

ನನ್ನ ಹೆಸರು ರಂಜನ್

► Where do you live?

ನೀನು/ ನೀವು ಎಲ್ಲಿ ವಾಸವಾಗಿದ್ದೀಯ/ವಾಸವಾಗಿದ್ದೀರಿ?

Neenu/Neevu elli Vaasavaagiddeeya/Ri?

ನೀನು/ನೀವು ಎಲ್ಲಿ ವಾಸವಾಗಿದ್ದೀಯ/ರಿ

► I live near Ghantaghar

ನಾನು ಘಂಟಘರ್‌ನ ಬಳಿ ವಾಸವಾಗಿದ್ದೇನೆ

Naanu Gantagarna Bali Vaasa Vaagiddene

ನಾನು ಘಂಟಘರ್‌ನ ಬಳಿ ವಾಸವಾಗಿದ್ದೇನೆ

► How old are you?

ನಿಮ್ಮ ವಯಸ್ಸು ಎಷ್ಟು?

Nimma Vayassu Eshtu?

ನಿಮ್ಮ ವಯಸ್ಸು ಎಷ್ಟು?

► That building is tall

ಆ ಕಟ್ಟಡ ಎತ್ತರವಾಗಿದೆ

Aa Kattada etharavaagidhe

ಆ ಕಟ್ಟಡ ಎತ್ತರವಾಗಿದೆ.

► She is beautiful

ಅವಳು ಸುಂದರವಾಗಿದ್ದಾಳೆ

Avalu Sundaravaadiddaale

ಅವಳು ಸುಂದರವಾಗಿದ್ದಾಳೆ

► I like Bengali sweets

ನಾನು ಬಂಗಾಲಿ ಸಿಹಿತಿಂಡಿಯನ್ನು ಇಷ್ಟಪಡುತ್ತೇನೆ.

Naanu Bangaali Thindiyannu ishtapaduthene

ನಾನು ಬಂಗಾಲಿ ಸಿಹಿ ತಿಂಡಿಯನ್ನು ಇಷ್ಟಪಡುತ್ತೇನೆ

► I love birds

ನಾನು ಪಕ್ಷಿಗಳನ್ನು ಇಷ್ಟಪಡುತ್ತೇನೆ

Naanu Pakshigalannu Ishtapaduthene

ನಾನು ಪಕ್ಷಿಗಳನ್ನು ಇಷ್ಟಪಡುತ್ತೇನೆ

► Where is Railway station?

ರೈಲ್ವೆ ನಿಲ್ದಾಣ ಎಲ್ಲಿದೆ?(ಉಗಿಬಂಡಿ ನಿಲ್ದಾಣ ಎಲ್ಲಿದೆ?)

Railve NiladhaanaEllidhe?(ugibandi Niladaana ellidhe?)

ರೈಲ್ವೆ ನಿಲ್ದಾಣ ಎಲ್ಲಿದೆ? (ಉಗಿಬಂಡಿ ನಿಲ್ದಾಣ ಎಲ್ಲಿದೆ?)

► How far is the Bus Terminal from here?

ಬಸ್ಸಿನ ಕೊನೆಯ ನಿಲ್ದಾಣ ಇಲ್ಲಿಂದ ಎಷ್ಟು ದೂರವಿದೆ?

Bassina koneya niladhaana illimdha Eshtu dhooravidhe?

ಬಸ್ಸಿನ ಕೊನೆಯ ನಿಲ್ದಾಣ ಇಲ್ಲಿಂದ ಎಷ್ಟು ದೂರವಿದೆ?

► How long will it take to reach the Airport?

ವಿಮಾನ ನಿಲ್ದಾಣವನ್ನು ತಲುಪಲು ಎಷ್ಟು ಸಮಯ ಬೇಕಾಗುತ್ತದೆ?

Vimaana Niladhaanavannu thalupalu Eshtu samaya Bekaaguthadhe?

ವಿಮಾನ ನಿಲ್ದಾಣವನ್ನು ತಲುಪಲು ಎಷ್ಟು ಸಮಯ ಬೇಕಾಗುತ್ತದೆ?

► Is Mr. Raghunath there?

ಅಲ್ಲಿ ಶ್ರೀ ರಘುನಾಥ್ ಇದ್ದಾರೆಯೇ?

Alli Shree Raghunaath idhaareye?

ಅಲ್ಲಿ ಶ್ರೀ ರಘುನಾಥ್ ಇದ್ದಾರೆಯೇ?

► Please tell him to call back as soon as he is free

ಅವನಿಗೆ/ಅವರಿಗೆ ಬಿಡುವಾದ ತಕ್ಷಣ ಕರೆ ಮಾಡಲು ದಯವಿಟ್ಟು ಹೇಳು/ಹೇಳು

Avanige/Avarige biduvaadha thakshana kare maadalu Hlu/Heli

ಅವನಿಗೆ/ಅವರಿಗೆ ಬಿಡುವಾದ ತಕ್ಷಣ ಕರೆ ಮಾಡಲು ದಯವಿಟ್ಟು ಹೇಳು/ಹೇಳು

► How much will it cost?

ಇದರ ಬೆಲೆ ಎಷ್ಟು?

Idhara bele Eshtu?

ಇದರ ಬೆಲೆ ಎಷ್ಟು?

► Excuse me

ನನ್ನನ್ನು ಕ್ಷಮಿಸಿ

Nannannu Kshamisi

ನನ್ನನ್ನು ಕ್ಷಮಿಸಿ



- From which Platform can I get the train for Chandigarh?

ಯಾವ ಪ್ಲಾಟ್ ಫಾರಂನಿಂದ ನನಗೆ ಚಂಡಿಗರ್ಹ್‌ಗೆ ಹೋಗಲು ರೈಲು ಸಿಗುತ್ತದೆ?

Yava plat formnimdha nanage Changarge Hogalu railu siguthadheye?

ಯಾವ ಪ್ಲಾಟ್ ಫಾರ್ಮ್ ನಿಂದ ನನಗೆ ಚಂಡಿಗರ್ಹ್ ಗೆ ಹೋಗಲು ರೈಲು ಸಿಗುತ್ತದೆ?

- Does this train stop at Aligarh?

ಈ ರೈಲು ಅಲಿಗರ್ಹ್‌ನಲ್ಲಿ ನಿಲ್ಲುತ್ತದೆಯೇ?

Ee railu Aligharadhalli nilluthadheye?

ಈ ರೈಲು ಅಲಿಗರ್ಹ್‌ನಲ್ಲಿ ನಿಲ್ಲುತ್ತದೆಯೇ?

- How many kids do you have?

ನಿನ್ನಿಗೆ/ನಿನ್ನಿಗೆ ಎಷ್ಟು ಮಂದಿ ಮಕ್ಕಳು ಇದ್ದಾರೆ?

Nimage/Ninage Eshtu Mandhi makkalu iddare?

ನಿನ್ನಿಗೆ/ನಿನ್ನಿಗೆ ಎಷ್ಟು ಮಂದಿ ಮಕ್ಕಳು ಇದ್ದಾರೆ?

- This gift is wonderful

ಈ ಉಡುಗೊರೆ ಅದ್ಭುತವಾಗಿದೆ.

E udugore adhbuthavaagidhe.

ಈ ಉಡುಗೊರೆ ಅದ್ಭುತವಾಗಿದೆ.

- It is really pretty

ಅದು ನಿಜವಾಗಲೂ ಸುಂದರವಾಗಿದೆ

Adhu Nijavaagalu Sundharavaagidhe.

ಅದು ನಿಜವಾಗಲೂ ಸುಂದರವಾಗಿದೆ.

- Food is delicious

ಊಟ/ತಿಂಡಿ ರುಚಿಕರವಾಗಿದೆ

Uta/Thindi Ruchikaravaagidhe

ಊಟ ತಿಂಡಿ ರುಚಿಕರ ವಾಗಿದೆ

- Congratulations

ಅಭಿನಂದನೆಗಳು

Abhinandhanegalu

ಅಭಿನಂದನೆಗಳು

- You look lovely

ನೀನು ಮುದ್ದಾಗಿ ಕಾಣಿಸುತ್ತಿದ್ದೀಯ/ ನೀವು ಮುದ್ದಾಗಿ

ಕಾಣಿಸುತ್ತಿದ್ದೀರ

Neenu Mudhaagi Kaanisuthiddeeya/

Kaanisuthiddeera

ನೀನು ಮುದ್ದಾಗಿ ಕಾಣಿಸುತ್ತಿದ್ದೀಯ/ನೀವು ಮುದ್ದಾಗಿ ಕಾಣಿಸುತ್ತಿದ್ದೀರ

- Wish you happy new year

ನಿಮ್ಮಿಗೆ/ನಿನ್ನಿಗೆ ಹೊಸವರ್ಷದ ಶುಭಾಶಯಗಳು

Nimage/Ninage Hosa varshadha

Shubhaashayagalu

ನಿಮ್ಮಿಗೆ/ನಿನ್ನಿಗೆ ಹೊಸ ವರ್ಷದ ಶುಭಾಶಯಗಳು

- I wish you all the happiness

ನಾನು ನಿಮ್ಮಿಗೆ ಪೂರ್ಣಸಂತೋಷವನ್ನು ಬಯಸುತ್ತೇನೆ

Naanu Nimage Poornasanthoshavannu bayasuthene.

ನಾನು ನಿಮ್ಮಿಗೆ ಪೂರ್ಣ ಸಂತೋಷವನ್ನು ಬಯಸುತ್ತೇನೆ

- Congratulations on your marriage

ನಿಮ್ಮ/ನಿನ್ನ ಮದುವೆಗೆ ಅಭಿನಂದನೆಗಳು

Nimma/Ninna Madhuvege Abhinandhanegalu

ನಿಮ್ಮ/ನಿನ್ನ ಮದುವೆಗೆ ಅಭಿನಂದನೆಗಳು

- Keep your eyes wide open before marriage and half- shut afterwards

ನಿನ್ನ ಕಣ್ಣುಗಳನ್ನು ಮದುವೆಗೆ ಮುನ್ನ ಪೂರ್ಣವಾಗಿಯೂ ಅನಂತರ ಅರ್ಧವೂ ತೆರೆದಿಟ್ಟುಕೊ ನಿನ್ನ ಕಣ್ಣುಗಳನ್ನು ಮದುವೆಗೆ ಮುನ್ನ ಪೂರ್ಣವಾಗಿಯೂ ಅನಂತರ ಅರ್ಧವೂ ತೆರೆದಿಟ್ಟುಕೊಳ್ಳಿ

Nimma Kannugalannu Madhuvege Munna poorthiyaagiyo Ananthara Ardhavoo theredhittukolli

ನಿನ್ನ ಕಣ್ಣುಗಳನ್ನು ಮದುವೆಗೆ ಮುನ್ನ ಪೂರ್ಣವಾಗಿಯೂ ಅನಂತರ ಅರ್ಧವೂ ತೆರೆದಿಟ್ಟುಕೊ ನಿನ್ನ ಕಣ್ಣುಗಳನ್ನು ಮದುವೆಗೆ ಮುನ್ನ ಪೂರ್ಣವಾಗಿಯೂ ಅನಂತರ ಅರ್ಧವೂ ತೆರೆದಿಟ್ಟುಕೊಳ್ಳಿ

(Courtesy : Shri C. V. Srinath Sastry
Directorate of Information Technology
Government of Karnataka,
Bangalore (Karnataka)
Phone: 080-5279611(O), 6645865(R)
E-mail : sastry@ead.cmmacs.ernet.in)