INDIC TEXT SEGMENTATION

Editor	Swaran Lata, DeitY
	Somnath chandra, DeitY
	Manoj Kumar Jain, DeitY
	Prashant Verma, WSI
Date	
This Version	
Latest Version	
Latest Proposed Update	
Revision	

Summary

The Unicode Standard UAX#29 is generally based text elements: user-perceived characters, words, and sentences. This report covers the guidelines for determination of boundaries of Indic characters.

This report also proposes the new ABNF valid text segmentation definition called Orthographic Indic Syllable , which defines the possible types of Indic text boundaries.

Some of the properties that based on text segmentation i.e First Initial drop letter ,Verticle layout , Line breaking etc. w.r.t Indian languages have been covered in this report.

Status

The document is prepared under WSI(Web Standardization Intiative) of TDIL programme of Department of Electronics and Information Technology , Govt. of India and in consultation with various experts and Linguists from Institutions/Industries.This is the first report based on Indic Layout requirements : W3C first public working draft.

Contents

---

1 Orthographic syllable (Ref section 3 of UAX#29)

An Orthographic syllable includes Independent vowel or a base consonant and/or any combination of the following characters in the text stream:

- Consonant/s and consonant + virama sequences
- vowel signs
- Modifiers

The above definition of Orthographic syllable is based on the tailored grapheme cluster disscussed in section 3 of UAX#29 report.The possible rules , which are required for tailoring grapheme cluster boundaries of Indic syllable defined in the following section 3 of this report.

2. Indic grapheme cluster boundaries (Ref section 3 of UAX#29)

A string of Unicode-encoded text often needs to be broken up into text elements programmatically. Common examples of text elements include what users think of as characters, words, lines (more precisely, where line breaks are allowed), and sentences. The precise determination of text elements may vary according to orthographic conventions for a given script or language.

Indian languages syllable boundaries should be based on tailored Grapheme Cluster Boundaries to conform Indic orthographic syllable definition.

Grapheme cluster boundaries are important for collation, regular expressions, UI interactions (such as mouse selection, arrow key movement, backspacing), segmentation for vertical text, identification of boundaries for first-letter styling.Grapheme clusters should be tailored to meet the Indic Orthographic Syllable requirements.

2.1 Sample tailored Grapheme Cluster Boundaries(Ref table 1a of UAX#29)

Some of the additional examples of Indic syllable based on tailored grapheme cluster boundaries and Indic syllable definition(see section 3) are shown in the following table , which is reffered from table 1a of UAX#29.

क्या	0915 (क) DEVANAGARI LETTER KA 094D (्) DEVANAGARI SIGN VIRAMA 092F (य) DEVANAGARI LETTER SSA	Devanagari kya

	093E (ा) DEVANAGARI SIGN AA	
स्थि	0938 (स) DEVANAGARI LETTER SA 094D (्) DEVANAGARI SIGN VIRAMA 0925 (थ) DEVANAGARI LETTER THA 091C (ि) DEVANAGARI LETTER I	Devanagari sthi
स्तः	0938 (स) DEVANAGARI LETTER SA 0924 (त) DEVANAGARI LETTER TA 0903 (ः) DEVANAGARI Sign Visarga	Devanagari sth
त्क्ल	0924 (त) DEVANAGARI LETTER TA 094D (्) DEVANAGARI SIGN VIRAMA 0915 (क) DEVANAGARI LETTER KA 094D (्) DEVANAGARI SIGN VIRAMA 0932 (ल) DEVANAGARI LETTER LA	Devanagari tkl

3 Indic orthographic syllable boudary rules : ABNF Valid segmentation-Proposed solution for layout issues in Indian languages

ABNF Valid Segmentation based Indic orthographic syllable definition is provided here for correct and standardized representation of Indian languages layout. This will address various issues mentioned in the following sections. This definition will be useful in order to get the uniform display of Indic layout in the browsers, applications, Digital publishing etc.

3.1 ABNF Valid segmentation-Proposed solution for layout issues in Indian languages

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] | {CH}C[v][m] | CH

The linguistic definition of Indic orthographic 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 which may or may not include a single nukta

v (lower case) is any dependent vowel or vowel sign [sign [V_{vs} has been used as symbol in Unicode for dependent vowel of full vowel V e.g **AA_{vs}**]

H is Virama

| is a rule separator

[ ] - The enclosed items is optional under this bracket

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

3.2. Indic syllable boundary determination(Ref section 8 of UAX#29)

The no break rules of Indian languages syllable based on orthographic syllable definition are shown in the below table as discussed in the section 8 of UAX#29 report.

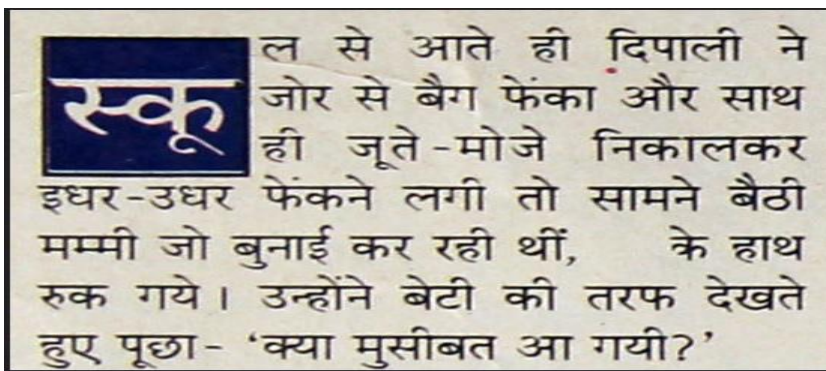| Rules	Do not break between
V[m]	Independent vowel and Modifier
{CH}C[v][m]	one or more consonant + virama sequences and Consonant zero or more consonant + virama sequences , Consonant and dependent vowel sign zero or more consonant + virama sequences , Consonant and modifier zero or more consonant + virama sequences, Consonant ,dependent vowel sign and modifier
CH	consonant with virama (applicable only for those Indian languages where pure consonant appears at the end of the word)

The precise list of characters with their Unicode code points of all the categories i.e C, H etc of Indic syllable definition are enclosed as Appendix 1.

4. Text Segmentation(Ref section 3 of UAX#29)

According to the section 3 of UAX#29 , Grapheme cluster are important for identification of bounderies of verticle text, first letter styling, the below section discuss about the boundaries of the same.

4.1 Initial letter styling

The initial letter styling in Indic scripts must be selected on the basis of orthographic syllables, rather than individual letter forms. Here is some of the examples of Initial letter styling based on the tailored grapheme cluster.
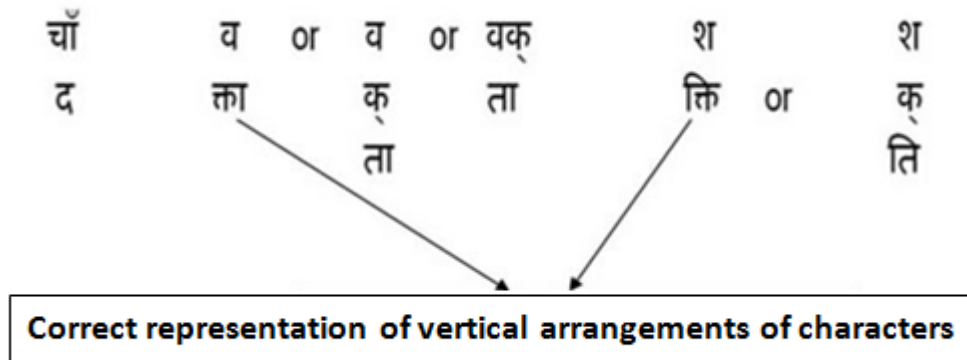

0938: स DEVANAGARI LETTER SA

094D: ॒ DEVANAGARI SIGN VIRAMA

0925: थ DEVANAGARI LETTER THA

093F: ि DEVANAGARI VOWEL SIGN I

syllable

स्थिति

4.2   Verticle segmentation

In vertical arrangement of characters writing each character on a new line may not be suitable in Indian languages. The vertical arrangements of characters are sometimes used in Indian texts. In order to form correct arrangements, it is preferred to follow tailored grapheme cluster approach. Variations of vertical arrangement of the characters in Hindi is represent below :

Variations in vertical arrangements

Correct representation of vertical arrangements of characters

4.3   Line boundaries

When inline-level content is laid out into lines, it is broken across line boxes. Such a break is called a line break. In most writing systems, in the absence of hyphenation a line break occurs only at word boundaries. Many writing systems use spaces or punctuation to explicitly separate words, and line break opportunities can be identified by these characters. Line breaking, also known as word wrapping, is the process of breaking a section of text into lines such that it will fit in the available width of a page, window or other display area.

4.3.1   Guiding principles of Line breaking for Indian languages

In Indic writing system , it is preferred that line breaks at word boundaries ,if required following principles may be adhered :

Rule 1: New line cannot begin with following symbols/Punctuation marks. Also these should be retain with the associated text :

Symbols	Character name	Unicode code-point
l	DEVANAGARI DANDA	U + 0964
ll	DEVANAGARI DOUBLE DANDA	U + 0965
)	RIGHT PARENTHESIS	U + 0029
+	PLUS SIGN	U + 002B
*	ASTERISK	U + 002A
‐	HYPHENATIONPOINT-VISIBLE HYPHEN HYPHENATION-SOFT HYPHEN	U + 2027

		U+ 00AD
/	SOLIDUS	U + 002F
,	COMMA	U + 002C
.	FULL STOP	U + 002E
:	COLON	U + 003A
;	SEMICOLON	U + 003B
=	EQUALS SIGN	U + 003D
>	GREATER–THAN SIGN	U + 003E
]	RIGHT SQUARE BRACKET	U + 005D
_	LOW LINE	U + 005F
|	VERTICAL LINE	U + 007C
}	RIGHT CURLY BRACKET	U + 007D
~	TILDE	U + 007E
%	PERCENT SIGN	U + 0025

Rule 2: The definition of Indic orthographic syllable may be used to break the line and a hyphen should be at the breaking point so that word can be read intuitively.

However the language specific morpho-phonemic rules and industry practices (from media, publishing and grammar books) could be used for hyphenation. U+ 00AD (soft hyphen) is used in some languages such as Tamil and Malayalam.

Rule 3: The hyphenated words can be broken at the hyphenation point (U + 2027) e.g.:

- नर-नारी should be treated as:
- नर- on the first line and नारी on the next line

Rule 4: Expression with mathematical symbol should be treated as single unit so that at the end of the line expression should not breaks at operator level

Rule 5: Breaking should not be allowed at numerical values such as currency values, year etc. e.g.

"100.00" or "10,000", nor in "12:59"

References

- 1. http://w3c.github.io/ilreq/
- 2. http://www.unicode.org/reports/tr29/tr29-26.html

Appendix 1 :

S.No.	Scripts	Consonant(C)	Vowel Independent	Vowel dependent v	Modifier	Halant(Virama)

			v			
1.	Devanagari	A) DEVANAGARI LETTER KA...DEVANAGARI LETTER HA (0915..0939) B) DEVANAGARI LETTER QA....DEVANAGARI LETTER YYA. (0958..095F) C) DEVANAGARI LETTER MARWARI DDA...DEVANAGARI LETTER BBA (0978..097F)	A) DEVANAGARI LETTER SHORT A...DEVANAGARI LETTER AU (0904..0914) B) DEVANAGARI LETTER VOCALIC RR...DEVANAGARI LETTER VOCALIC LL (0960..0961) C) DEVANAGARI LETTER CANDRA A...DEVANAGARI LETTER UUE (0972..0977)	A) DEVANAGARI VOWEL SIGN OE (093A) B) DEVANAGARI VOWEL SIGN OOE (093B) C) DEVANAGARI VOWEL SIGN AA...DEVANAGARI VOWEL SIGN II (093E...0940) D) DEVANAGARI VOWEL SIGN U...DEVANAGARI VOWEL SIGN AI (0941..0948) E) DEVANAGARI VOWEL SIGN CANDRA O...DEVANAGARI VOWEL SIGN AU(0949..094C) F) DEVANAGARI VOWEL SIGN PRISHTHAMATRA E...DEVANAGARI VOWEL SIGN AW (094E...094F) G) DEVANAGARI VOWEL SIGN CANDRA LONG E...DEVANAGARI VOWEL SIGN UUE (0955..0957) H) DEVANAGARI VOWEL SIGN VOCALIC L...DEVANAGARI VOWEL SIGN VOCALIC LL (0962..0963)	A) DEVANAGARI SIGN INVERTED CANDRABINDU...DEVA NAGARI SIGN ANUSVARA (0900...0902) B) DEVANAGARI SIGN VISARGA (0903) C) DEVANAGARI SIGN AVAGRAHA (093D) D) DEVANAGARI SIGN NUKTA (093C)	A)DEVANAGARI SIGN VIRAMA (094D)
2.	Bengali & Assamese	A) BENGALI LETTER KA...BENGALI LETTER NA (0995..09A8) B) BENGALI LETTER PA...BENGALI LETTER RA (09AA.. 09B0) C) BENGALI LETTER LA (09B2) D) BENGALI LETTER SHA...BENGALI LETTER HA (09B6...09B9) E) BENGALI LETTER KHANDA TA (09CE) F) BENGALI LETTER RRA..BENGALI LETTER RHA (09DC...09DD) G)BENGALI LETTER YYA (09DF) H) BENGALI LETTER RA WITH MIDDLE DIAGONAL...BENGALI LETTER RA WITH LOWER DIAGONAL (09F0...09F1)	A) BENGALI LETTER A...BENGALI LETTER VOCALIC L (0985...098C) B) BENGALI LETTER E...BENGALI LETTER AI (098F...0990) C) BENGALI LETTER O...BENGALI LETTER AU (0993...0994) D) BENGALI LETTER VOCALIC RR...BENGALI LETTER VOCALIC LL (09E0...09E1)	A) BENGALI VOWEL SIGN AA...BENGALI VOWEL SIGN II(09BE..09C0) B) BENGALI VOWEL SIGN U..BENGALI VOWEL SIGN VOCALIC RR (09C1..09C4) C) BENGALI VOWEL SIGN E...BENGALI VOWEL SIGN AI (09C7...09C8) D) BENGALI VOWEL SIGN O..BENGALI VOWEL SIGN AU (09CB...09CC) E) BENGALI AU LENGTH MARK (09D7) F) BENGALI VOWEL SIGN VOCALIC L...BENGALI VOWEL SIGN VOCALIC LL(09E2...09E3)	A) BENGALI ANJI = siddham, siddhirastu (0980) B) BENGALI SIGN CANDRABINDU (0981) C) BENGALI SIGN ANUSVARA (0982) D) BENGALI SIGN VISARGA (0983) E) BENGALI SIGN AVAGRAHA (09BD) F)BENGALI SIGN NUKTA(09BC)	A) BENGALI SIGN VIRAMA (09CD)
3.	Gurmukhi	A) GURMUKHI LETTER KA...GURMUKHI LETTER NA (0A15...0A28) B) GURMUKHI LETTER PA...GURMUKHI LETTER RA (0A2A...0A30) C) GURMUKHI LETTER LA...GURMUKHI LETTER LLA (0A32...0A33) D) GURMUKHI LETTER VA...GURMUKHI LETTER SHA (0A35...0A36) E) GURMUKHI LETTER SA...GURMUKHI LETTER HA (0A38..0A39) F) GURMUKHI LETTER KHHA...GURMUKHI LETTER	A) GURMUKHI LETTER A...GURMUKHI LETTER UU (0A05...0A0A) B) GURMUKHI LETTER EE...GURMUKHI LETTER AI (0A0F...0A10) C) GURMUKHI LETTER OO...GURMUKHI LETTER AU (0A13...0A14) D) GURMUKHI	A) GURMUKHI VOWEL SIGN AA...GURMUKHI VOWEL SIGN II (0A3E...0A40) B) GURMUKHI VOWEL SIGN U...GURMUKHI VOWEL SIGN UU (0A41...0A42) C) GURMUKHI VOWEL SIGN EE...GURMUKHI VOWEL SIGN AI (0A47...0A48) D) GURMUKHI VOWEL SIGN OO...GURMUKHI VOWEL SIGN AU (0A4B...0A4C)	A) GURMUKHI SIGN ADAK BINDI...GURMUKHI SIGN BINDI (0A01...0A02) B) GURMUKHI TIPPI (A70) C) GURMUKHI SIGN VISARGA (0A03) D) GURMUKHI SIGN NUKTA (A3C) E) GURMUKHI TIPPI (0A70) E) GURMUKHI ADDAK (A71)	A) GURMUKHI SIGN VIRAMA (0A4D)

		RRA (0A59...0A5C) G) GURMUKHI LETTER FA (0A5E) GURMUKHI SIGN YAKASH (0A75)	IRI..GURMUKHI URA (0A72..0A73)			
4.	Gujarati	A) GUJARATI LETTER KA...GUJARATI LETTER NA (0A95...0AA8) B) GUJARATI LETTER PA...GUJARATI LETTER RA (0AAA...0AB0) C) GUJARATI LETTER LA...GUJARATI LETTER LLA (0AB2...0AB3) D) GUJARATI LETTER VA...GUJARATI LETTER HA (0AB5...0AB9) E) GUJARATI LETTER ZHA (0AF9)	A) GUJARATI LETTER A...GUJARATI VOWEL CANDRA E (0A85...0A8D) B) GUJARATI LETTER E...GUJARATI VOWEL CANDRA O (0A8F..0A91) C) GUJARATI LETTER O...GUJARATI LETTER AU (0A93...0A94) D) GUJARATI LETTER VOCALIC RR...GUJARATI LETTER VOCALIC LL (0AE0...0AE1)	A) GUJARATI VOWEL SIGN AA...GUJARATI VOWEL SIGN II (0ABE...0AC0) B) GUJARATI VOWEL SIGN U...GUJARATI VOWEL SIGN CANDRA E (0AC1...0AC5) C) GUJARATI VOWEL SIGN E...GUJARATI VOWEL SIGN AI (0AC7..0AC8) D) GUJARATI VOWEL SIGN CANDRA O (0AC9) E) GUJARATI VOWEL SIGN O...GUJARATI VOWEL SIGN AU (0ACB...0ACC) F) GUJARATI VOWEL SIGN VOCALIC L...GUJARATI VOWEL SIGN VOCALIC LL (0AE2..0AE3)	A) AGUJARATI SIGN CANDRABINDU...GUJARATI SIGN ANUSVARA (0A81...0A82) B) GUJARATI SIGN VISARGA (0A83) C) GUJARATI SIGN AVAGRAHA (0ABD) D)GUJARATI SIGN NUKTA (0ABC)	A) GUJARATI SIGN VIRAMA (ACD)
5.	Odia	A) ODIA LETTER KA...ODIA LETTER NA (0B15...0B28) B) ODIA LETTER PA...ODIA LETTER RA (0B2A...0B30) C) ODIA LETTER LA...ODIA LETTER LLA (0B32...0B33) D) ODIA LETTER VA...ODIA LETTER HA (0B35...0B39) E) ODIA LETTER RRA...ODIA LETTER RHA (0B5C...0B5D) F) ODIA LETTER YYA (0B5F) G) ODIA LETTER WA (0B71)	A) ODIA LETTER A...ODIA LETTER VOCALIC L (0B05...0B0C) B) ODIA LETTER E...ODIA LETTER AI (0B0F...0B10) C) ODIA LETTER O...ODIA LETTER AU(0B13...0B14) D) ODIA LETTER VOCALIC RR...ODIA LETTER VOCALIC LL (0B60..0B61)	A) ODIA VOWEL SIGN AA (0B3E) B) ODIA VOWEL SIGN I(0B3F) C) ODIA VOWEL SIGN II(0B40) D) ODIA VOWEL SIGN U...ODIA VOWEL SIGN VOCALIC RR (0B41...0B44) E) ODIA VOWEL SIGN E...ODIA VOWEL SIGN AI (0B47...0B48) F) ODIA VOWEL SIGN O...ODIA VOWEL SIGN AU (0B4B...0B4C) G) ODIA AI LENGTH MARK (0B56) H) ODIA AU LENGTH MARK (0B57) I) ODIA VOWEL SIGN VOCALIC L...ODIA VOWEL SIGN VOCALIC LL (0B62...0B63)	A)ODIA SIGN CANDRABINDU (0B01) B) ODIA SIGN ANUSVARA (0B02) C) ODIA SIGN VISARGA (0B03) D) ODIA SIGN AVAGRAHA (0B3D) E) ODIA SIGN NUKTA (0B3C)	A) ODIA SIGN VIRAMA (0B4D)
6.	Tamil	A)TAMIL LETTER KA (0B95) B) TAMIL LETTER NGA...TAMIL LETTER CA(0B99...0B9A) C) TAMIL LETTER JA (0B9C) D) TAMIL LETTER NYA...TAMIL LETTER TTA (0B9E...0B9F) E) TAMIL LETTER NNA...TAMIL LETTER TA (0BA3...0BA4) F) TAMIL LETTER NA...TAMIL LETTER PA (0BA8...0BAA) G) TAMIL LETTER MA...TAMIL LETTER HA (0BAE...0BB9)	A) TAMIL LETTER A...TAMIL LETTER UU (0B85..0B8A) B) TAMIL LETTER E..TAMIL LETTER AI (0B8E...0B90) C) TAMIL LETTER O...TAMIL LETTER AU (0B92...0B94)	A) TAMIL VOWEL SIGN AA...TAMIL VOWEL SIGN I (0BBE...0BBF) B) TAMIL VOWEL SIGN II (0BC0) C) TAMIL VOWEL SIGN U...TAMIL VOWEL SIGN UU (0BC1...0BC2) D) TAMIL VOWEL SIGN E...TAMIL VOWEL SIGN AI (0BC6...0BC8) E) TAMIL VOWEL SIGN O...TAMIL VOWEL SIGN AU (0BCA...0BCC)	A) TAMIL SIGN ANUSVARA (0B82) B) TAMIL SIGN VISARGA (0B83)	A) TAMIL SIGN VIRAMA (0BCD)
7.	Telugu	A) TELUGU LETTER KA...TELUGU LETTER NA	A) TELUGU LETTER	A) TELUGU VOWEL SIGN AA...TELUGU VOWEL SIGN II	A)TELUGU SIGN COMBINING	A)TELUGU SIGN VIRAMA (0C4D)

No.	Language					
		(0C15...0C28) B) TELUGU LETTER PA...TELUGU LETTER HA (0C2A...0C39)	A...TELUGU LETTER VOCALIC L (0C05...0C0C) B) TELUGU LETTER E...TELUGU LETTER AI (0C0E...0C10) C) TELUGU LETTER O...TELUGU LETTER AU (0C12...0C14) D) TELUGU LETTER VOCALIC RR...TELUGU LETTER VOCALIC LL (0C60...0C61)	(0C3E...0C40) B) TELUGU VOWEL SIGN U...TELUGU VOWEL SIGN VOCALIC RR (0C41...0C44) C) TELUGU VOWEL SIGN E...TELUGU VOWEL SIGN AI (0C46...0C48) D) TELUGU VOWEL SIGN O...TELUGU VOWEL SIGN AU (0C4A...0C4C) E) TELUGU LENGTH MARK...TEL UGU AI LENGTH MARK (0C55...0C56) F) TELUGU VOWEL SIGN VOCALIC L...TELUGU VOWEL SIGN VOCALIC LL (0C62...0C63)	CANDRABINDU ABOVE (0C00) B) TELUGU SIGN CANDRABINDU...TELU GU SIGN ANUSVARA (0C01...0C02) C) TELUGU SIGN VISARGA (0C03) D) TELUGU SIGN AVAGRAHA (0C3D)	
8.	Kannada	A) KANNADA LETTER KA...KANNADA LETTER NA (0C95...0CA8) B) KANNADA LETTER PA...KANNADA LETTER LLA (0CAA...0CB3) C) KANNADA LETTER VA...KANNADA LETTER HA (0CB5...0CB9) D) KANNADA LETTER FA (0CDE)	A) KANNADA LETTER A...KANNADA LETTER VOCALIC L (0C85...0C8C) B) KANNADA LETTER E...KANNADA LETTER AI (0C8E...0C90) C) KANNADA LETTER O...KANNADA LETTER AU(0C92...0C94) D)KANNADA LETTER VOCALIC RR...KANNADA LETTER VOCALIC LL (0CE0...0CE1)	A) KANNADA VOWEL SIGN AA (0CBE) B) KANNADA VOWEL SIGN I (0CBF) C) KANNADA VOWEL SIGN II...KANNADA VOWEL SIGN VOCALIC RR (0CC0..0CC4) D)KANNADA VOWEL SIGN E (0CC6) E) KANNADA VOWEL SIGN EE...KANNADA VOWEL SIGN AI (0CC7...0CC8) F) KANNADA VOWEL SIGN O...KANNADA VOWEL SIGN OO (0CCA..0CCB) G)KANNADA VOWEL SIGN AU (0CCC) H) KANNADA LENGTH MARK...KANNADA AI LENGTH MARK (0CD5...0CD6) I) KANNADA VOWEL SIGN VOCALIC L...KANNADA VOWEL SIGN VOCALIC LL (0CE2...0CE3)	A)KANNADA SIGN CANDRABINDU (0C81) B) KANNADA SIGN ANUSVARA (0C82) C) KANNADA SIGN VISARGA(0C83) D) KANNADA SIGN AVAGRAHA(0CBD) E) KANNADA SIGN NUKTA(0CBC) F) ꭓKANNADA SIGN JIHVAMULIYA(CF1) G) ∞ KANNADA SIGN UPADHMANIYA (CF2)	A) KANNADA SIGN VIRAMA(0CCD)
9.	Malayalam	A) MALAYALAM LETTER KA...MALAYALAM LETTER TTTA (0D15...0D3A) MALAYALAM LETTER CHILLU NN..MALAYALAM LETTER CHILLU K (0D7A..0D7F)	A) MALAYALAM LETTER A...MALAYALAM LETTER VOCALIC L (0D05...0D0C) B) MALAYALAM LETTER E...MALAYALAM LETTER AI (0D0E..0D10) C) MALAYALAM LETTER O...MALAYALAM LETTER AU (0D12...0D14) D) MALAYALAM LETTER ARCHAIC II (0D5F) E) MALAYALAM LETTER VOCALIC RR... MALAYALAM LETTER	A) MALAYALAM VOWEL SIGN AA...MALAYALAM VOWEL SIGN II (0D3E...0D40) B) MALAYALAM VOWEL SIGN U...MALAYALAM VOWEL SIGN VOCALIC RR (0D41...0D44) C) MALAYALAM VOWEL SIGN E..MALAYALAM VOWEL SIGN AI (0D46...0D48) D) MALAYALAM VOWEL SIGN O...MALAYALAM VOWEL SIGN AU (0D4A..0D4C) E) MALAYALAM AU LENGTH MARK (0D57) F) MALAYALAM VOWEL SIGN VOCALIC L...MALAYALAM VOWEL SIGN VOCALIC LL (0D62...0D63)	A) MALAYALAM SIGN CANDRABINDU (0D01) B) MALAYALAM SIGN ANUSVARA (0D02) C) MALAYALAM SIGN VISARGA(0D03) D)MALAYALAM SIGN AVAGRAHA (0D3D)	A)MALAYALAM SIGN VIRAMA (0D4D)

			VOCALIC LL(0D60...0D61)			
10.	Meetei Mayek	A) MEETEI MAYEK LETTER CHA..MEETEI MAYEK LETTER SSA (AAE2...AAEA) B) MEETEI MAYEK LETTER KOK...MEETEI MAYEK LETTER HUK (ABC0...ABCD) C) MEETEI MAYEK LETTER PHAM (ABD0) D) MEETEI MAYEK LETTER GOK..MEETEI MAYEK LETTER BHAM (ABD2...ABDA) E) MEETEI MAYEK LETTER UN = u (ABCE) F) MEETEI MAYEK LETTER I (ABCF) G) MEETEI MAYEK LETTER ATIYA = a (ABD1) H) MEETEI MAYEK LETTER NA LONSUM = n (ABDF) I) MEETEI MAYEK LETTER TIL LONSUM = t (ABE0) J) MEETEI MAYEK LETTER NGOU LONSUM = ng (ABE1) K) MEETEI MAYEK LETTER I LONSUM = i(ABE2)	A) MEETEI MAYEK LETTER E...MEETEI MAYEK LETTER O (AAE0...AAE1) B) MEETEI MAYEK LETTER UN...MEETEI MAYEK LETTER I (ABCE...ABCF) C) MEETEI MAYEK LETTER ATIYA (ABD1)	A) MEETEI MAYEK VOWEL SIGN II (AAEB) B) MEETEI MAYEK VOWEL SIGN UU...MEETEI MAYEK VOWEL SIGN AAI (AAEC...AAED) C) MEETEI MAYEK VOWEL SIGN AU...MEETEI MAYEK VOWEL SIGN AAU (AAEE...AAEF) D) MEETEI MAYEK VOWEL SIGN ONAP...MEETEI MAYEK VOWEL SIGN INAP (ABE3..ABE4) E) MEETEI MAYEK VOWEL SIGN ANAP (ABE5) F) MEETEI MAYEK VOWEL SIGN YENAP...MEETEI MAYEK VOWEL SIGN SOUNAP (ABE6...ABE7) G) MEETEI MAYEK VOWEL SIGN UNAP (ABE8) H) MEETEI MAYEK VOWEL SIGN CHEINAP...MEETEI MAYEK VOWEL SIGN NUNG (ABE9...ABEA)	A)MEETEI MAYEK VOWEL SIGN VISARGA (AAF5)	A) MEETEI MAYEK APUN IYEK (ABED)